Agglomeration and the spatial determinants of productivity and trade

Anthony Briant

ÉCOLE DES PONTS PARISTECH

École Doctorale
ÉCONOMIE PANTHÉON-SORBONNE

Laboratoire
PARIS-JOURDAN SCIENCES ÉCONOMIQUES
UMR 8545 CNRS-EHESS-ENPC-ENS

Déterminants de la productivité et du commerce: le rôle de la proximité géographique

Thèse pour obtenir le grade de
Docteur de l'École des Ponts ParisTech
en Sciences Économiques

Présentée et soutenue publiquement par

Anthony BRIANT

le 16 Avril 2010

Directeur de thèse : Pierre-Philippe COMBES
Co-Directrice de thèse : Miren LAFOURCADE

Jury :

Dominique BUREAU,	Ingénieur général des Ponts et Chaussées, Professeur chargé de cours à l'Ecole Polytechnique
Pierre-Philippe COMBES,	Directeur de recherche au CNRS, Université d'Aix-Marseille
Miren LAFOURCADE,	Professeur des Universités, Université de Paris-Sud XI
Philippe MARTIN,	Professeur des Universités, Institut d'Études Politiques de Paris (rapporteur)
Henry OVERMAN,	Reader, Department of Geography and Environment, London School of Economics
Sébastien ROUX,	Administrateur INSEE, Centre de Recherche en Economie et Statistique de l'INSEE
William STRANGE,	RioCan Real Estate Investment Trust Professor of Real Estate and Urban Economics, Rotman School of Management, University of Toronto (rapporteur)

Remerciements

Je tiens tout d'abord à remercier Pierre-Philippe Combes, mon directeur de thèse, et Miren Lafourcade, ma co-directrice de thèse. Après avoir encadré mon mémoire de DEA, ils ont accepté de diriger ma thèse. Deux chapitres de cette thèse sont coécrits avec eux. Ces articles ont été l'occasion de comprendre et d'apprendre, par l'exemple, les efforts et la rigueur que nécessitent l'écriture d'un article de recherche. Je les en remercie vivement et espère sincèrement que nous aurons l'occasion de collaborer sur de nouveaux articles dans le futur. Au-delà de la qualité de leur encadrement, Pierre-Philippe et Miren ont tous deux des qualités humaines qui rendent la collaboration facile et agréable.

Cette thèse doit également beaucoup à mes autres co-auteurs juniors: Yoann Barbesol, Muriel Barlet et Laure Crusson. Confronter nos idées et partager nos lectures m'ont beaucoup aidé à avancer dans ma compréhension des phénomènes économiques étudiés dans cette thèse.

J'ai passé, durant cette thèse, presque deux ans au Département des Études Économiques d'Ensemble (D3E), à l'INSEE, afin de pouvoir avoir accès aux données d'entreprises. Je tiens particulièrement à remercier Didier Blanchet, Sébastien Roux et Pauline Givord pour m'avoir offert cette opportunité. Les membres de ce département m'ont également beaucoup apporté tant par leur connaissance fine des sources statistiques que par la qualité de leurs remarques sur mes travaux. Sans les citer tous, je tiens à remercier plus particulièrement Simon Quantin, Muriel Roger, Xavier Boutin, Claude Picart, Patrick Sillard, et Brigitte Rigot.

Je tiens à adresser des remerciements tout particulier à Thierry Mayer et Gilles Duranton. À plusieurs reprises, Thierry Mayer a pris le temps de lire et d'émettre un avis critique éclairant sur les sujets que je traitais. Gilles Duranton m'a offert la possibilité de passer un an au Département d'Économie de l'Université de Toronto. Cette visite et nos échanges durant cette période m'ont permis de mieux comprendre quelles sont les exigences de la recherche en économie à un niveau international.

J'ai également une pensée amicale pour mes collègues de PSE, et plus particulièrement Florian Mayneris, Julie Rochut, Laurent Gobillon, Luc Arrondel et Andrew Clark.

Enfin, je remercie Bernard Caillaud pour ses conseils avisés, ainsi que Marie-Christine Paoletti, Claude Tu, Barbara Chahnamian, Isabelle Lelièvre pour avoir facilité, à maintes reprises, mes démarches administratives et mon orientation dans un montage institutionnel parfois complexe.

Mes derniers remerciements iront à mon épouse et à ma fille. Muriel a été d'un soutien sans faille et d'une patience remarquable au cours de cette thèse. Elle a su partager mes moments de doute et a accepté de nombreuses concessions pour me permettre de mener mes travaux de recherche dans de bonnes conditions. Margot a, depuis bientôt 8 mois, la faculté de faire disparaître les tracas du quotidien par ses sourires et ses éclats de rire...

Introduction en français

L'existence et la pérennité des villes sont la preuve indéniable de la tendance naturelle de l'activité humaine à se concentrer spatialement. Ce phénomène de concentration spatiale s'observe également pour des secteurs d'activité particuliers. L'étude de l'organisation spatiale des secteurs d'activité, et plus particulièrement la détection des schémas de concentration spatiale, sont un sujet d'étude ancien pour les économistes, depuis les travaux d'Alfred Marshall (1890) au moins.

Pour autant, l'agglomération de l'activité humaine génère de la congestion, fait croître le prix des facteurs de production immobiles et, éventuellement, accroît le degré de concurrence locale. Plusieurs questions sont donc d'intérêt dans cette littérature: la concentration spatiale est-elle propre à quelques secteurs d'activité particuliers ou bien un phénomène partagé par le plus grand nombre ? Quels sont les bénéfices que les entreprises tirent de cette concentration spatiale ? Dans quelle mesure ces bénéfices font-ils plus que compenser les coûts que l'agglomération ne manque pas de générer ? Quelle est l'étendue géographique de ces externalités ? L'ensemble de ces questions forme la toile de fond de cette thèse.

Sur les origines de la concentration spatiale

Le point de départ de l'analyse des phénomènes d'agglomération est le *théorème d'impossibilité spatiale*, prouvé par Starrett (1978). Celui-ci énonce que si l'espace est homogène et qu'il n'existe pas d'invisibilités ou des rendements croissants, alors tout équilibre concurrentiel en présence de coûts de transport doit se caractériser par une ensemble de localisation en autarcie, où chaque bien est produit à petite échelle (voir Ottaviano et Thisse, 2004, pour un commentaire détaillé).

A contrario, il est possible d'observer des zones où l'activité est agglomérée dès lors que l'espace est hétérogène ou bien qu'il existe une forme ou une autre d'indivisibilités ou de rendements croissants, sous l'hypothèse que les coûts de transport sont non nuls.[1] La notion d'espace hétérogène renvoie à l'idée que chaque territoire bénéficie de dotations spécifiques (dotations naturelles, technologies ou aménités) qui, au regard des dotations des autres territoires, favorisent le développement d'un type particulier d'activités. Il s'agit là des avantages comparatifs analysés dans les théories traditionnelles du commerce international. Ce raisonnement est bien sûr valable au niveau mondial, mais peut

[1] Voir Combes, Mayer et Thisse (2008b, chapitre 2) pour une présentation détaillée de la prise en compte de l'espace dans la pensée économique.

également être invoqué à une niveau géographique plus réduit. Certains territoires, au sein même d'un pays, peuvent bénéficier de dotations qui favorisent l'installation d'un secteur d'activité particulier, comme les houillères dans le Nord Est de la France. Cependant, lorsque l'échelle géographique considérée est plus réduite, il est plus probable que l'espace soit homogène. Ainsi la théorie des avantages comparatifs, même quand ceux-ci sont définis dans une acception large, ne suffit plus à expliquer l'ensemble des phénomènes de concentration observés au niveau d'un pays par exemple (voir pour une illustration, Ellison et Glaeser, 1999, pour les États-Unis).

Dans cette situation, des explications alternatives à l'agglomération de l'activité économique doivent être mobilisées. En présence de coûts à l'échange positif, l'agglomération, suivant Starrett (1978), s'explique par l'existence de rendements croissants, ou d'*économies d'agglomération*. Ces économies existent dès que la productivité d'un individu s'accroît lorsqu'il ou elle se trouve à proximité d'autres individus. Les économies d'agglomération peuvent être des *externalités pures*, comme par exemple des externalités de connaissance. Ces économies d'agglomération peuvent également transiter par le marché. Si un producteur et un fournisseur sont plus proches géographiquement, il est possible qu'ils deviennent plus productifs, car la proximité élimine un certain nombre des coûts de transaction liés à l'éloignement. Il n'y a pas dans ce cas d'externalités manifestes (voir Glaeser, 2008).

Les rendements croissants peuvent être internes à l'entreprise. Krugman (1991) (et la littérature qui a suivi en Économie Géographique) a ainsi montré que les entreprises pouvaient avoir tendance à se concentrer spatialement quand les coûts de transport diminuent. Dans ce schéma, les entreprises gagnent à se localiser dans un marché plus large, de sorte à pouvoir exploiter au maximum les rendements croissants de leur technologie de production (voir Combes, Mayer et Thisse, 2008b, pour une revue exhaustive de la littérature). Au contraire, l'Économie Urbaine traditionnelle suppose l'existence d'externalités de production entre des entreprises produisant avec des rendements constants (voir Henderson, 1974). Ainsi, les rendements croissants en jeu sont externes à l'entreprise. Dans ce schéma, les entreprises bénéficient aussi d'un marché local plus large ou de la proximité d'entreprises exerçant dans le même secteur d'activité suivant les types d'externalités en jeu. Cette thèse s'intéresse, avant tout, à ce second type de rendements croissants.

Quand les économies d'agglomération ne sont pas cantonnées aux entreprises d'un secteur donné, elles permettent l'agglomération de l'ensemble des secteurs et donnent naissance aux villes. Elles sont alors nommées *économies d'urbanisation*. Au contraire, quand ces économies ne concernent que les entreprises d'un secteur donné, elles donnent naissance à des pôles industriels spécialisés et sont nommées *économies de localisation*. Les économies d'urbanisation, comme de localisation, peuvent également se comprendre comme une manière de réduire les coûts de déplacement des biens, des hommes et des idées (voir Glaeser, 2008), et, par conséquent, accroissent la productivité individuelle, des travailleurs et des entreprises. Mettre en évidence l'existence d'économies d'agglomération et évaluer leur ampleur sont des questions fondamentales de la littérature en économie urbaine et régionale, puisque, sans réponse, nous ne saurions expliquer l'existence des villes. Le premier chapitre de cette thèse fournit une preuve de l'existence des économies d'agglomération, en comparant les schémas de concentration spatiale dans les secteurs de services et dans les secteurs manufacturiers en France. Dans les chapitres 2 et 3, nous évaluons les gains de productivité des entreprises françaises liés aux économies d'urbanisation et de localisation.

Les sources de ces économies d'agglomération sont également un sujet largement débattu dans la littérature. Plusieurs mécanismes ont été proposés pour expliquer les rendements croissants donnant naissance aux villes ou aux pôles industriels. Duranton et Puga (2004) proposent une revue exhaustive de la littérature sur le sujet et classent les modèles suivant trois types: les modèles basés sur le partage, l'appariement et l'apprentissage. Une industrie de bien final plus concentrée spatialement attire un plus grand nombre de fournisseurs produisant des produits différentiés. Dans le même ordre d'idées, un marché du travail plus large permet ainsi le développement d'un plus grand nombre de tâches et des gains liés à une plus grande spécialisation. L'appariement entre les employeurs et les employés est également supposé être plus facile et de meilleure qualité dans un marché du travail plus dense. Enfin, la proximité géographique entre entreprises facilite la création, la diffusion et l'accumulation de connaissances. Dans le chapitre 4 de cette thèse, nous fournissons une preuve indirecte de l'accumulation de connaissances lorsque les agents économiques se concentrent. Nous étudions comment la concentration spatiale des immigrés entre départements français influence le commerce international de ces mêmes départements vers le pays d'origine de ces immigrés.

Les économies d'urbanisation et de localisation sont par définition localisées, limitées à un petit espace géographique. La diffusion spatiale des économies d'agglomération est un dernier axe important de recherche, comme l'ont récemment souligné Rosenthal et Strange (2004). Du fait de contraintes liées aux données, les chercheurs font souvent l'hypothèse que ces économies s'inscrivent dans des territoires définis administrativement. Cependant, mener des analyses empiriques suivant un découpage géographique arbitraire peut avoir des conséquences importantes sur les résultats trouvés. Le chapitre 5 conclut cette thèse en étudiant la sensibilité des exercices économétriques développés dans les chapitres précédents à un changement dans le découpage géographique utilisé. Autrement dit, nous étudions comment la taille et la forme des unités spatiales influencent les mesures de la concentration spatiale, de l'ampleur des économies d'agglomération ou bien des déterminants spatiaux du commerce.

La concentration spatiale : première preuve de l'existence d'économies d'agglomération

Mesurer une concentration spatiale excessive d'un secteur d'activité est la première preuve de l'existence au sein de ce secteur d'économies d'agglomération. Mesurer la concentration spatiale consiste à décrire les inégalités spatiales en terme de production ou d'emploi. Les économistes et les géographes ont développés un certain nombre d'outils permettant de rendre compte de ces inégalités spatiales (voir Combes, Mayer et Thisse, 2008b, chapitre 10).

Une première approche pour mesurer la concentration spatiale d'un secteur consiste à comparer la distribution spatiale de son emploi à la distribution spatiale de l'emploi total. Les mesures traditionnelles de concentration spatiale, comme l'indice de Gini, reposent sur cette méthode. Cependant, Ellison et Glaeser (1997) soulignent le fait qu'au sein d'un secteur, l'emploi est réparti entre un nombre limité d'établissements. Ceci introduit une forme de "granulosité" qui empêche que l'emploi sectoriel soit distribué de manière identique à l'emploi total. L'intuition est simple : même si les établissements d'un secteur étaient distribués de manière parfaitement aléatoire dans l'espace, la distribution spatiale de l'emploi sectoriel, du fait de cette granulosité, ne pourrait être parfaitement similaire à la distribution spatiale de l'emploi total. Ils proposent donc un indice de concentration spatiale qui corrige de la concentration industrielle de chaque secteur, c'est-à-dire du nombre d'établissements et de la distribution d'emploi entre établissements. Leur indice remplit au moins trois des six critères listés par Combes et Overman (2004) pour définir un indice idéal de concentration spatiale : l'indice est défini en comparaison à une distribution spatiale de référence bien établie, la significativité statistique de la concentration spatiale peut être évaluée, l'indice est comparable d'un secteur à l'autre. Néanmoins, leur indice repose sur un découpage géographique donné du territoire. L'indice est donc potentiellement sensible à la taille, la forme et la position relative des unités spatiales qui composent ce découpage. Ces problèmes sont dénommés Problèmes des Unités Spatiales Modifiables[2]. L'indice suggéré par Ellison et Glaeser (1997) échoue donc devant deux critères: être insensible à un changement dans le découpage géographique et être comparable d'une zone à l'autre[3].

Pour répondre à ces deux limites, une seconde approche consiste à travailler en espace continu. L'idée, initiée par Duranton et Overman (2005), est de considérer la densité des distances bilatérales entre paires d'établissements au sein d'un secteur d'activité. Ils testent si la densité observée est proche ou non d'une densité prédite dans le cas où les établissements du secteur seraient redistribués de manière aléatoire sur le territoire. Ils évaluent la significativité statistique de l'écart à l'hypothèse de distribution aléatoire en construisant un intervalle de confiance global autour de cette densité prédite[4]. Les distances bilatérales entre paires d'établissements sont calculées comme la distance à vol d'oiseau entre leurs coordonnées géographiques. Cette méthode est donc intensive en manipulation de données géo-localisées. L'indice proposé par Duranton et Overman (2005) remplit l'ensemble des propriétés listées par Combes et Overman (2004), à l'exception de la sensibilité à un changement de nomenclature industrielle.

Dans le chapitre 1, nous proposons une nouvelle méthode pour tester la concentration spatiale en espace continu. Nous montrons tout d'abord que la méthode proposée par Duranton et Overman (2005) dépend, de manière implicite, de la structure industrielle de

[2] Nous décrivons en détail ce Problème des Unités Spatiales Modifiables dans le chapitre 5.

[3] L'indice d'Ellison et Glaeser est également sensible à un changement de nomenclature industrielle.

[4] Voir l'annexe 1.7 du chapitre 1 pour une présentation plus formalisée de ces concepts.

chaque secteur, à savoir du nombre d'établissements et de la répartition de l'emploi entre établissements. Ainsi, leur approche n'est pas parfaitement adaptée à la comparaison intersectorielle. Nous suggérons donc une approche alternative qui repose aussi sur la densité des distances bilatérales entre établissements au sein d'un secteur. Nous construisons notre test de concentration spatiale sur la base d'une mesure de divergence entre fonctions de densité. L'idée est d'estimer la divergence entre la densité observée des distances au sein du secteur et la densité des distances entre l'ensemble des établissements de l'économie. Cette mesure n'est a priori pas comparable entre secteurs pour les mêmes raisons que celles avancées par Ellison et Glaeser (1997). C'est la raison pour laquelle nous nous inspirons de leur méthode pour rendre cette mesure de divergence parfaitement comparable d'un secteur à l'autre. Notre démarche reposant sur les distributions de distances, et non de l'emploi, nous sommes en mesure de donner une information sur l'étendue spatiale des phénomènes de concentration dans chaque secteur. Nous distinguons donc les secteurs où cette concentration intervient principalement à très courte distance (moins de 4 km), à moyenne distance (entre 4 et 40 km) et enfin à longue distance (entre 40 et 140 km). On peut ainsi proposer un tri des secteurs en fonction de la distance spécifique à laquelle ils apparaissent concentrés. Notre intuition est que cette distance dépend de manière prononcée du type d'externalités d'agglomération sous-jacent. Ainsi les secteurs dans lesquels les contacts face-à-face et les spillovers technologiques sont importants devraient se concentrer dans une rayon spatial limité. Au contraire, lorsque les forces d'agglomération sont liés au marché de l'emploi ou à la proximité de fournisseurs, cette concentration spatiale peut s'inscrire dans un espace géographique plus large.

On applique ensuite cette nouvelle méthodologie à la comparaison de l'organisation spatiale des secteurs de services aux entreprises et aux secteurs manufacturiers en France. Nous montrons tout d'abord que les services divergent plus souvent de la distribution aléatoire que les secteurs manufacturiers. Ensuite, nous montrons qu'une majorité des secteurs de services qui sont concentrés spatialement, le sont à courte distance. Autrement dit, ces secteurs s'organisent dans un petit nombre de pôles industriels spécialisés. Ceci est en accord avec l'intuition que certains de ces services aux entreprises ne se localisent que dans le coeur des plus grandes villes.

Évaluer l'ampleur des économies d'agglomération

La concentration spatiale excessive de l'activité économique est un premier signe de l'existence des économies d'agglomération. Mais quels sont les bénéfices que les entreprises tirent de cette agglomération ? La question est donc de quantifier l'ampleur des économies d'agglomération. Puga (2009) et Strange (2009) distinguent trois approches dans la littérature.

La première consiste à comparer la croissance de l'emploi entre villes ou entre pôles industriels. L'intuition est simple : si les entreprises sont plus productives dans les villes (ou dans les pôles industriels spécialisés), l'emploi doit y croître plus vite. Les papiers initiaux traitant de cette question, Glaeser, Kallal, Scheinkman et Schleifer (1992) et Henderson, Kuncoro et Turner (1995), relient la croissance à long terme de l'emploi sectoriel dans les villes américaines au degré de spécialisation locale (économies de localisation) et à la diversité du tissu économique local (économies d'urbanisation). Glaeser et al. (1992) concluent à une prédominance des économies d'urbanisation en montrant que la diversité du tissu économique local est positivement corrélée avec la croissance de l'emploi sectoriel. Au contraire, prenant en compte des effets dynamiques, Henderson et al. (1995) montrent que les effets de la spécialisation locale sont les plus importants. Pour la France, Combes (2000) et Combes, Magnac et Robin (2004) étudient la croissance à long terme de l'emploi dans des zones géographiques plus petites, les zones d'emploi[5] et distinguent la croissance de l'emploi dans les établissements existants (marge intensive) de l'installation de nouveaux établissements (marge extensive). Ils montrent que les établissements préalablement existants croient plus vite dans les zones avec un nombre important d'établissements de tailles différentiés, alors que les nouveaux établissements sont plus généralement localisés dans des zones avec un petit nombre d'établissements de tailles différentiées. Ces papiers soulignent également le caractère dynamique des économies d'agglomération. Ainsi Combes et al. (2004) montrent qu'en France les économies d'agglomération jouent à court terme et sont donc plutôt statiques. Au contraire, Henderson (1997) trouvent un effet retardé des économies d'agglomération sur la croissance de l'emploi local, avec un retard de 6 à 7 ans. Ces études sur la croissance de l'emploi local reposent sur des hypothèses fortes, notamment que toute croissance de la productivité se traduit par un accroissement de l'emploi, ce qui n'est pas toujours le cas (voir Combes et al., 2004, pour plus de détails). Aussi les équations de croissance en emploi local ne sont-elles pas forcément l'approche la plus adaptée pour évaluer l'ampleur des économies d'agglomération.

La seconde approche consiste à étudier les différentiels de rendement des facteurs de production - travail et capital foncier - entre villes ou pôles industriels. Si les entreprises gagnent à s'agglomérer, elles seront prêtes à attirer les travailleurs en leur offrant des salaires (nominaux) plus élevés et à payer plus cher leur terrain. Strange (2009) offre une revue de littérature sélective sur l' *urban wage premium* (la prime salariale urbaine), à savoir l'impact de l'agglomération (et plus particulièrement d'une plus forte densité en emploi) sur les salaires. Un contribution séminale est celle de Glaeser et Maré (2001) qui montrent que les travailleurs dans les villes de plus de 500 000 habitants aux États-Unis gagnent, en moyenne, des salaires 33% plus élevés que les travailleurs des zones rurales. Cette prime obtenue dans les grandes villes n'est plus que de l'ordre de 5 à 11% lorsque que l'on contrôle par un nombre important de caractéristiques individuelles, et que l'on s'efforce donc de comparer des individus identiques. En effet, l'avantage des données sur les salaires est de fournir au côté de l'information salariale, un ensemble de caractéris-

[5]Ces zones d'emploi sont des unités spatiales construites sur un critère économique précis par l'Institut National de la Statistique et des Études Économiques, comme l'entité géographique qui minimise les déplacements pendulaires domicile-travail transfrontaliers. Autrement dit, ce sont des zones dans lesquels la plupart des actifs vivent et travaillent.

tiques concernant les salariés[6] ce qui permet aux chercheurs de corriger un certain nombre de biais éventuels. Utilisant des données longitudinales sur un échantillon représentatif des salariés français, Combes, Duranton et Gobillon (2008) montrent que près de la moitié des disparités salariales entre les grandes villes et les campagnes s'explique par un tri spatial en fonction des qualifications. Autrement dit, les grandes villes attirent les travailleurs les plus qualifiés. Ce tri spatial biaise toute estimation naïve de la prime salariale urbaine, ce que Combes, Duranton, Gobillon et Roux (à paraître) nomment le biais de "qualité endogène du facteur travail". Combes et al. (2008) développent une méthodologie économétrique complexe qui leur permet de soustraire aux données brutes de salaires l'effet de caractéristiques individuelles observables et inobservables. Les disparités résiduelles entre salaires nets sont ensuite expliquées par des variables d'urbanisation et de localisation. Ils montrent ainsi que les variables d'urbanisation (plus précisément la densité en emploi) influencent fortement les salaires nets, alors que les variables de localisation, bien que positivement corrélées, n'ont qu'un impact économique faible. Combes et al. (à paraître) considère également une autre source de biais, celui de "quantité endogène de travail". Cette source de biais est due à la simultanéité entre la détermination des salaires (ou plus généralement de la productivité) et la densité. En effet, les zones les plus productives attirent plus de gens, et deviennent ainsi plus denses. Dans ce cas, il y a un causalité circulaire entre productivité et agglomération. Pour rompre cette causalité circulaire, Combes et al. (à paraître) utilisent des instruments historiques et géologiques, valides sous l'hypothèse qu'ils influencent les choix contemporains de localisation des agents mais pas leur productivité. Contrairement au "biais de qualité endogène du travail", le biais de simultanéité (ou causalité inverse) est de faible amplitude. Rosenthal et Strange (2008) développent également une stratégie d'identification par variables instrumentales sur des données de salaires américaines, et conclut aussi à un biais faible. La stratégie d'estimation des économies d'agglomération à partir des données de salaires doit toutefois être considérée avec précaution. En effet, dans un contexte d'équilibre spatial comme celui développé par Rosen (1979) et Roback (1982), les salaires sont déterminés de manière endogène par la migration des travailleurs. Les forces agissant sur les salaires sont multiples. Schématiquement, si les entreprises bénéficient d'externalités de production dans les villes, alors elles poussent les salaires à la hausse pour y attirer des travailleurs. Par ailleurs, si les travailleurs bénéficient dans les villes de certains avantages (aménités urbaines), ils seront alors prêts à accepter des salaires moindres pour y résider. Au final, les salaires résultent donc de l'interaction entre l'effet des externalités de production et des aménités de consommation (voir Glaeser, 2008, pour plus de détails sur le concept d'équilibre spatial).

La stratégie la plus directe pour quantifier les économies d'agglomération est d'en chercher l'effet sur la productivité des entreprises. Les premières tentatives dans ce sens ont cherché à lier une mesure locale agrégée de la productivité (par exemple un PIB par tête) à la taille du marché local (Moomaw, 1981 ; Nakamura, 1985 ; Henderson, 1986 ;

[6]Les papiers considérant l'impact de l'agglomération sur le coût du capital foncier sont plus rares, notamment du fait d'un manque d'informations pertinentes sur les caractéristiques de ce capital.

Ciccone et Hall, 1996 ; Ciccone, 2002). Cependant, une telle approche macro-économique se heurte aux problèmes d'identification évoqués plus haut, notamment à celui lié à une répartition spatiale des travailleurs qualifiés non aléatoire. C'est la raison pour laquelle les chercheurs ont rapidement utilisé des données individuelles d'entreprises pour mettre en évidence les effets des économies d'agglomération sur leur productivité. Les données individuelles permettent en effet de contrôler d'un certain nombre des déterminants individuels et sectoriels du processus de production. Henderson (2003) est le premier à introduire dans une fonction de production au niveau des établissements des variables d'agglomération. Ses données sont constituées d'un panel non-exhaustif d'établissements américains, observés tous les cinq ans, dans les secteurs des machines-outils et les secteurs high-tech. Ils disposent pour chaque établissement d'une information sur la valeur ajoutée, le capital et l'emploi. Il met en évidence l'existence d'économies de localisation dans les secteurs high-tech mais pas les secteurs de machines-outils. En effet, dans les secteurs high-tech, les firmes profitent à s'installer dans une zone où d'autres établissements exercent la même activité économique. Il montre aussi que les entreprises mono-établissements sont plus à même de générer et de profiter des externalités d'agglomération que les entreprises pluri-établissements.

Dans le chapitre 2, nous quantifions l'ampleur des économies d'agglomération sur la productivité des entreprises françaises, à l'aide de données individuelles détaillées issues des déclarations fiscales. Suivant Combes et al. (à paraître), nous utilisons une méthodologie en deux étapes. Dans une première étape, nous estimons une fonction de production du type Cobb-Douglas, dont les résidus constituent nos mesures de productivité individuelle. Dans un second temps, nous expliquons les disparités spatiales de productivité individuelle moyenne par des indicateurs pour les économies d'urbanisation et de localisation. Dans la première étape, la richesse des données individuelles à disposition nous permet de soustraire de la mesure de productivité l'effet d'un nombre important de ces déterminants individuels et sectoriels, non liés aux économies d'agglomération mais susceptibles de biaiser nos estimations. Nous contrôlons notamment de la qualité de la main d'oeuvre dans chaque établissement. Cela nous permet de nous prémunir d'un biais liés à la répartition géographique non aléatoire des travailleurs en fonction de leur qualification, évoquée plus haut. Nous soustrayons à cette mesure de productivité individuelle tout déterminant sectoriel. Là encore, certains secteurs, a priori plus productifs, peuvent avoir une tendance à se localiser dans les zones les plus denses. Dans ce cas, l'organisation spatiale non aléatoire de chaque secteur pourrait créer une corrélation positive entre agglomération et productivité, sans lien avec les effets que nous cherchons à mesurer. Dans la seconde étape, nous montrons que les entreprises localisées dans des territoires à forte densité en emploi (par exemple dans le 9^e décile pour la distribution de densité en emploi) sont, en moyenne, 8% plus productives que des entreprises comparables situées dans des territoires à faible densité en emploi (par exemple dans le 1^{er} décile pour la distribution de densité en emploi). Cet effet est économiquement important lorsqu'on le compare aux 2,2% de croissance de productivité moyenne annuelle enregistrée par les entreprises françaises sur la période 1993-1999. Nous mettons également en évidence un effet bénéfique de l'accessibilité d'un territoire au reste du marché national pour la productivité des entreprises qui y sont implantées. Par contre, une fois contrôlé de la densité en emploi, la diversité du tissu économique local ne semble joué que marginalement, et plutôt négativement, sur la productivité des entreprises. Concernant les économies de localisation, nous trouvons que les entreprises situées dans les territoires les plus spécialisés (ceux du 9^e décile pour la variable de spécialisation) sont, en moyenne, 5% plus productives que celles localisées dans les zones les moins spécialisées (à savoir celles dans le 1^{er} décile pour la variable de spécialisation). L'impact de la spécialisation locale est donc moins marqué que celui de la densité en emploi mais reste économiquement important. Nous mettons également en exergue une corrélation positive entre le niveau de qualification des travailleurs dans un territoire et la productivité des entreprises qui y sont implantées. Mais cette variable n'a pas de pouvoir explicatif important une fois contrôlé de la qualité de la main d'oeuvre dans chaque entreprise, et de la densité en emploi locale. Aussi, on ne rejette pas l'existence d'externalités de capital humain, mais leur effet semble modéré une fois contrôlé des variables précédentes.

Des développements théoriques récents (voir Melitz et Ottaviano, 2008) suggèrent que les entreprises d'une même industrie peuvent ne pas être distribuées de manière aléatoire sur le territoire. Les entreprises les plus productives sont a priori plus enclines à survivre à la concurrence plus féroce à l'oeuvre dans les territoires les plus denses. Suivant cette logique, les territoires les plus denses accueilleraient les firmes les plus productives par sélection ou éviction des firmes les moins productives. Ce genre de mécanismes de sélection est susceptible de biaiser nos estimations. Pour corriger partiellement de ce phénomène, nous introduisons dans la deuxième étape des effets fixes régionaux qui contrôlent du niveau moyen de productivité dans chaque région. L'idée est alors de comparer des entreprises opérant dans des territoires différents de la même région. Si le mécanisme de sélection s'effectue au niveau régional, la comparaison intra-régionale est plus raisonnable. Les résultats précédents restent robustes à l'introduction de tels effets fixes.

Dans la seconde étape, nous expliquons les différences *moyennes* de productivité entre territoires par les variables d'agglomération. Cependant, au sein même d'une industries, les producteurs restent fortement hétérogènes, et le comportement du "producteur moyen" ne fournit qu'une information partielle sur l'ensemble de la distribution de productivité. Dans le chapitre 3, nous considérons que les économies d'agglomération influencent non seulement la productivité moyenne des entreprises, mais qu'elles induisent aussi des déformations plus complexes dans la distribution locale de productivité. Nous suggérons donc que l'effet des économies d'agglomération ne peut être parfaitement compris sans considérer cette hétérogénéité entre producteurs. Pour y parvenir, nous utilisons une méthode par régressions quantiles qui nous permet de paramétriser de manière parcimonieuse l'impact des économies d'urbanisation et de localisation à différents points de la distribution de productivité. La technique des régressions quantiles, introduites par Koenker et Bassett (1978), peut être utiliser pour caractériser les déformations de la distribution conditionnelle de productivité quand la valeur des régresseurs change. Les coefficients estimés à différents points de la distribution peuvent s'interpréter comme la réaction différentiée de la variable dépendante à un même incrément des régresseurs, à différents points de la distribution conditionnelle. Dans notre situation, nous évaluons comment les économies d'agglomération influencent la productivité d'entreprises situées à différents quantiles de la distribution conditionnelle, et pas seulement la moyenne de cette distribution. Deux résultats émergent de notre analyse. Les entreprises sont non seulement plus productives dans les territoires à forte densité en emploi, mais les économies d'urbanisation induites par cette forte densité profitent plus aux entreprises les plus productives. Les entreprises sont également en moyenne plus productives dans les zones les plus spécialisées, et, contrairement aux variables d'urbanisation, la spécalisation locale joue de manière uniforme sur les entreprises, quelque soit leur niveau de productivité. Ces deux résultats n'ont pas a priori d'interprétations évidentes dans la littérature théorique existante. En effet, le nombre d'études considérant l'hétérogénéité verticale entre entreprises reste restreint dans cette littérature. Combes, Duranton, Gobillon, Puga et Roux (2009) ont récemment offert une interprétation à l'impact différentié de la densité en emploi sur les entreprises. Dans leur modèle, les producteurs sont initialement hétérogènes, et les travailleurs sont d'autant plus productifs qu'ils sont embauchés par une entreprise productive, et cet effet est amplifié par les interactions dont ils sont susceptibles de bénéficier dans un territoire à forte densité. Dans ce modèle, l'hétérogénéité initiale est amplifiée dans les territoires les plus denses par l'intermédiaire des travailleurs. Malgré cette première piste d'analyse, il nous semble que les résultats mis en exergue dans cette étude suscite une réflexion théorique pour comprendre l'effet différentié des variables d'urbanisation et de localisation.

Sur l'importance de la diffusion locale d'information

Le chapitre 4 s'intéresse à des problématiques de commerce international. Cependant, deux points de convergence avec les thématiques évoquées précédemment peuvent être soulignés. Dans un premier temps, le partage local de connaissances est souvent évoqué dans la littérature théorique et empirique comme l'un des moteurs principaux du développement local (voir Lucas, 1988, 2001). Pour le dire simplement, le stock local de connaissances constitue un bien public local qui rend les entreprises plus productives. Partant de la même intuition, nous étudions dans le chapitre 4 comment la concentration spatiale des immigrés dans un territoire donné influence les performances commerciales de ce territoire. En effet, des économistes, notamment Rauch (2001), ont émis l'idée que les immigrés possédaient une information spécifique sur leur pays d'origine qui pouvait faciliter la création de liens commerciaux entre leur territoire d'accueil et ce pays d'origine. Ces connaissances particulières peuvent permettre de réduire les coûts informationnels qui empêche parfois aux opportunités d'échanges commerciaux de se réaliser. En effet, malgré la démocratisation et la diffusion rapide des moyens de communication à échelle planétaire, les coûts informationnels restent l'une des barrières principales aux flux commerciaux mondiaux, comme le soulignent by Anderson et Van Wincoop (2004).

Rauch (2001) évoque également deux autres dimensions suivant lesquelles les immigrés pourraient favoriser le développement commercial. La première est par l'existence, au sein des réseaux transnationaux, de mécanismes, souvent implicites, de sanctions et d'exclusions, qui évitent les pratiques illégales et se substituent à un cadre légal fragilisé dans le pays partenaire. Par ailleurs, les immigrés peuvent avoir une préférence particulière pour les biens produits dans leur pays d'origine, ce qui développerait les flux d'importations de leur région d'accueil. Finalement, pour toutes ces raisons, les réseaux transnationaux peuvent se substituer à la proximité pour réduire les coûts à l'échange. Ce chapitre s'intéresse donc plus aux économies de réseaux plus que les économies d'agglomération proprement dites, mais les points de convergence entre les deux sont nombreux.

Un second point de rapprochement de ce chapitre avec les chapitres précédents est d'ordre économétrique. Dans ce chapitre, nous étudions la question de la simultanéité entre les flux de commerce et d'immigration. Le rôle promoteur de commerce des immigrés est bien établi empiriquement au niveau des pays. Ainsi, Gould (1994), Head et Ries (1998) et Girma et Yu (2002) montrent une corrélation positive et significative entre la présence d'immigrés et les flux de commerce vers et depuis leur pays d'origine aux Etats-Unis, au Canada et en Grande-Bretagne. Cependant, à l'échelle nationale, il est à craindre que la corrélation mise en exergue ne soit pas causalité. En effet, on peut penser à un certain nombre de déterminants communs des flux de commerce et d'immigration qui, s'ils sont oubliés des régressions, peuvent être la source d'une corrélation positive trompeuse entre commerce et immigration. Il en va ainsi pour les liens coloniaux, le partage d'une langue ou d'une culture commune par exemple. La corrélation peut aussi être trompeuse s'il y a causalité inverse, c'est-à-dire que l'existence de flux commerciaux entre deux pays incitent les émigrés de l'un à immigrer dans l'autre. Du fait de ces sources de biais important, nous étudions dans ce chapitre le lien entre commerce et immigration à un niveau infranational, celui des départements français. Nous contrôlons des déterminants nationaux par l'introduction d'effets fixes propres à chaque pays. Les variations dans les flux d'exportations et d'importations des départements français mis en regard des différences dans les stock d'immigrés qu'ils accueillent permet d'estimer l'effet promoteur de commerce de ces immigrés. De manière à évacuer les différentes sources de biais évoquées plus haut, nous utilisons aussi une stratégie par variables instrumentales. Du fait d'une forme d'hystérésis dans les choix de localisation des populations étrangères par nationalité, les stocks retardés de population immigrée sont fortement corrélées au stock présent. Pour autant, ces stocks n'influencent certainement plus les flux contemporains de commerce. A ce titre, ces stocks retardés apparaissent comme de bons instruments. Nous montrons ainsi que l'immigration a un effet positif et significatif sur les flux de commerce. Doubler le nombre d'immigrés dans un département français accroit de 7% les exportations de ce département vers le pays d'origine de ces immigrés, et de 4% ses importations.

Nous poussons ensuite l'exercice plus loin en étudiant les effets différentiés de l'immigration suivant deux dimensions: la complexité du bien échangé, et la qualité des institutions dans le pays d'origine des immigrés. Le fait que les connaissances possédées par les immigrés puissent être plus valorisées pour le commerce de biens complexes est suggéré par Rauch et Trindade (2002). Ces auteurs montrent que les pays d'Asie du Sud-

Est qui accueillent la diaspora chinoise la plus importante commercent plus les uns avec les autres. Ils montrent aussi que l'effet promoteur de commerce de la diaspora chinoise est plus marqué pour les biens les plus différentiés, et donc pour lequel le coût informationnel à l'échange est le plus marqué. Par ailleurs, Anderson et Marcouiller (2002) et Berkowitz, Moenius, et Pistor (2006) montrent que la qualité des institutions est élément déterminant dans les volumes de commerce d'un pays. Berkowitz et al. (2006) montrent que l'importance des institutions est d'autant plus marquée dans le commerce de biens complexes, pour lesquels il est difficile de spécifier l'ensemble des caractéristiques dans un contrat. Il est donc possible que, par le jeu des sanctions et exclusions propres aux membres d'un réseau transnational, les réseaux d'immigration se substituent à la faiblesse des institutions, notamment lorsque les contrats de commerce sont difficile à écrire, c'est-à-dire lorsque les biens à échanger sont plus complexes. Nous montrons en effet que les immigrants ont un effet promoteur de commerce plus marqué sur l'importations de biens complexes, et ce quelque soit la qualité des institutions dans le pays partenaire. Par contre, pour les biens les moins complexes, l'effet des immigrants n'est significatif que lorsque ceux-ci sont originaires de pays avec une faible qualité des institutions. Les effets sont moins différentiés pour les exportations. Néanmoins, l'effet promoteur de commerce des immigrés est plus marqué pour ceux venant de pays avec des institutions faibles.

Quand la géographie s'en mêle : Le Problème des Unités Spatiales Modifiables

Le dernier chapitre de cette thèse apporte une contribution méthodologique. Nous nous concentrons sur le Problème des Unités Spatiales Modifiables évoqué plus haut. Nous avons souligné en présentant l'indicateur d'Ellison et Glaeser que celui-ci reposait sur la définition d'un espace discrétisé. Cet indicateur est donc potentiellement sensible à cette définition. Considérons, par exemple, une industrie équi-répartie entre deux zones. Cette industrie devrait apparaître plus concentrée si ces deux zones sont contigües, que si elles sont fortement éloignées. L'indicateur d'Ellison et Glaeser ne permet pas pour autant de distinguer ces deux situations. Eviter ce genre d'effets frontières est la raison première pour laquelle Duranton et Overman préfèrent travailler en espace continu. De la même manière, de récentes tentatives ont été faites pour s'abstraire d'un découpage géographique arbitraire dans l'évaluation de l'ampleur des économies d'agglomération et ainsi travailler en espace continu (voir Rosenthal et Strange, 2003, 2008).

Plus généralement, la plupart des travaux en économie régionale reposent sur des données initialement individuelles et localisées comme des points sur une carte. Ces points sont ensuite, pour une raison ou pour une autre, agrégées dans des unités spatiales de taille et de forme prédéfinies, comme les villes ou les régions. Agréger l'information ponctuelle en une information surfacique peut avoir des conséquences sur les capacités du chercheur à mesurer correctement les phénomènes économiques sous-jacents. La sensibilité des résultats statistiques aux choix d'un découpage géographique particulier est dénommé Problème des Unités Spatiales Modifiables. Dans le chapitre 5, nous évaluons comment le changement de taille (ou de manière équivalente, du nombre) des unités spatiales, et de leur forme (ou de manière équivalente, du dessin de leurs frontières) altère les résultats des exercices économétriques menés dans les chapitres précédents. En parallèle, nous comparons les distorsions induites par le Problème des Unités Spatiales Modifiables à celles induites par une mauvaise spécification des régressions.

Cet exercice est d'autant plus important que, dans la plupart des travaux empiriques, la maille géographique est contrainte par le disponibilité des données. Ainsi, de nombreux travaux ont évalué l'impact de l'agglomération sur la productivité des entreprises ou des travailleurs au niveau des pays, des régions européennes, des états américains, ou bien d'échelles spatiales plus petites comme les comtés américains ou bien les zones d'emploi françaises. Les effets mesurés diffèrent d'une étude à l'autre mais il est difficile de distinguer entre la part qui revient au changement d'unités spatiales et celle qui revient aux mécanismes économiques proprement dit.

Ainsi dans le chapitre 5, nous commençons par évaluer comment le degré de concentration spatiale varie entre trois types de zonages différents (un zonage administratif, un zonage basé sur un carroyage, et un zonage aléatoire). Nous comparons également ces différences à celles introduites, pour un zonage donné, par l'utilisation d'un indicateur de Gini plutôt qu'un indicateur d'Ellison et Glaeser. Nous poursuivons par des exercices économétriques. Nous estimons l'effet d'un changement de découpage géographique sur le lien entre productivité du travail et densité en emploi d'une part, et sur l'estimation d'équations gravitaires de commerce, d'autre part. Dans ce dernier cas, nous nous concentrons sur les coûts physiques et informationnels à l'échange évoqués au chapitre 4.

Tous ces exercices pointent vers la même conclusion : tant que le niveau d'agrégation n'est pas trop large, les effets du changement de taille jouent peu, et ceux liés au changement de forme jouent encore moins. Dans tous les cas, ces effets sont du second ordre au regard des biais introduits par une mauvaise spécification. Cependant, lorsque l'agrégation se fait à un niveau trop large, typiquement les régions en France, les résultats sont beaucoup moins robustes. Nous en dérivons quelques conseils pour travailler sur données spatiales. Une attention particulière doit être apportée par les chercheurs aux points suivants : 1 - la taille de la maille géographique au regard du processus ponctuel sous-jacent, 2 - la manière dont les données sont agrégées, par sommation ou moyenne, 3 - le degré d'autocorrélation spatiale dans les données initiales. Le Problème des Unités Spatiales Modifiables est, en effet, plus marqué quand les variables de part et d'autre d'une équation ne sont pas agrégées de la même manière. Dans les régressions gravitaires, par exemple, les flux de commerce, la variable dépendante, sont sommés, alors que les distances, une variable explicative, sont moyennées. Des trois exercices, ce dernier est le plus sensible au Problème des Unités Spatiales Modifiables.

Enfin, nous suggérons que si un maillage a été dessiné spécifiquement au regard d'un critère économique bien défini, comme les zones d'emploi en France, il doit être utilisé de préférence. Nous arrivons à une conclusion plutôt optimiste : tant que l'agrégation en s'effectue pas un niveau trop large, les problèmes liés au changement d'échelle et de forme restent d'ampleur limitée en comparaison des problèmes de spécification. Les chercheurs devraient donc mettre l'accent sur ce dernier aspect.

Déterminants de la productivité et du commerce : le rôle de la proximité géographique

Résumé long:

L'existence et la pérennité des villes est la preuve indéniable de la tendance naturelle de l'activité humaine à se concentrer spatialement. Ce phénomène de concentration spatiale s'observe également pour des entreprises exerçant leur activité dans une industrie particulière, donnant naissance à des pôles industriels. La mesure, les causes et les conséquences de cette concentration spatiale sont un sujet d'étude ancien chez les économistes, depuis les travaux d'Alfred Marshall (1890) au moins.

A une échelle infranationale, l'agglomération de l'activité économique, ou plus spécifiquement de certains secteurs d'activité, ne peut pas reposer uniquement sur l'existence d'avantages comparés exogènes propres à chaque territoire, tels qu'ils sont mobilisés dans les théories traditionnelles du commerce international. Les économistes ont donc étudié ces phénomènes en ayant recours à la notion d'externalités d'agglomération. Ces externalités existent dès qu'un agent voit sa productivité augmenter lorsqu'il est à proximité d'autres agents. Les gains économiques à l'agglomération sont alors endogènes. Plusieurs questions traversent la littérature quant à ces externalités d'agglomération: la concentration spatiale est-elle propre à quelques secteurs particuliers ou bien un phénomène partagé par le plus grand nombre ? Quels sont les bénéfices que les entreprises tirent de cette concentration spatiale ? Dans quelle mesure ces bénéfices font-ils plus que compenser les coûts que l'agglomération ne manque pas de générer ? Quelle est l'étendue géographique de ces externalités ? Comment cette concentration spatiale agit-elle sur d'autres grandeurs économiques, telles que les flux commerciaux ? L'ensemble de ces questions forment la toile de fond de cette thèse.

Dans le premier chapitre, nous développons une nouvelle méthodologie statistique permettant de rendre compte de la concentration spatiale d'un secteur d'activité en considérant le territoire national comme un espace continu. En effet, les méthodologies précédentes s'appuient pour la plupart sur un espace discrétisé, à savoir sur un ensemble d'entités géographiques prédéfinies. Or, il est possible que le choix d'un système particulier d'entités géographiques (zones d'emploi, départements, régions) influence la mesure de la concentration spatiale, ce que l'on nomme le Problème des Unités Spatiales Modifiables. On applique par ailleurs cette méthodologie pour comparer l'organisation spatiale des secteurs de services aux entreprises à celle des secteurs manufacturiers traditionnels en France. Cette comparaison met en exergue une tendance plus lourde des secteurs de services aux entreprises à se concentrer spatialement, et dans un périmètre géographique plus restreint. Ces résultats sont en accord avec l'intuition que ces activités requièrent des contacts face à face et un ensemble d'échanges informels facilités par la proximité géographique.

Dans les chapitres 2 et 3, nous évaluons l'ampleur des gains à l'agglomération en terme de productivité pour les entreprises. Nous évaluons l'impact relatif des externalités dites d'*urbanisation*, qui dépendent de la taille totale du marché local, à celui des externalités de *localisation*, qui dépendent quant à elles de la proximité aux entreprises exerçant leur

activité dans le même secteur. Dans le chapitre 2, nous montrons que la taille globale du marché local, mesurée par la densité en emploi, et sa relative spécialisation sont deux déterminants importants des différences spatiales de productivité moyenne entre entreprises. Nous avons bien entendu pris soin d'ôter de cette mesure de productivité l'ensemble de ses déterminants non liés aux choix de localisation mais qui pourraient biaiser nos estimations. Ainsi, les entreprises dans les zones d'activités les plus denses (i.e. celles du 9^e décile de la distribution de densité en emploi) sont, en moyenne, 8% plus productives que les entreprises dans les zones d'activité les moins denses (i.e. celles du 1^{er} décile de la distribution de densité en emploi). Ceci équivaut à quatre ou cinq années de croissance de la productivité (au regard des résultats de croissance enregistrés sur la décennie 1990). Les entreprises dans les zones les plus spécialisées (i.e. celles du 9^e décile de la distribution de la variable de spécialisation) sont aussi, en moyenne, 5% plus productives que celles dans les zones les moins spécialisées (i.e. celles du 1^{er} décile de la distribution de la variable de spécialisation).

Dans le chapitre 3, nous poussons l'exercice plus loin en se demandant si, derrière ce gain moyen, les résultats varient fortement d'un producteur à l'autre. Pour ce faire, nous utilisons les méthodes de régressions quantiles qui nous permettent d'évaluer l'impact de la densité en emploi et de la spécialisation locale sur des entreprises situées à différents quantiles de la distribution conditionnelle de productivité. Nous montrons que les firmes les plus productives sont celles qui bénéficient le plus des externalités d'urbanisation. Au contraire, les externalités de localisation semblent jouer de manière identique sur l'ensemble des entreprises, indépendamment de leur productivité.

Le chapitre 4 s'intéresse à un sujet connexe à ceux évoqués dans les chapitres précédents. Nous étudions dans quelle mesure la concentration spatiale des personnes immigrées influence le commerce international des départements qui les accueillent avec les pays d'origine de ces immigrés. Autrement dit, nous relions la distribution spatiale des immigrés entre départements français à la distribution spatiale des flux de commerce de ces mêmes départements avec les pays d'origine des immigrés. Nous montrons que le rôle des immigrants dans la création de commerce est important. Doubler les effectifs de personnes immigrées dans un département induit une croissance de 7%, en moyenne, de ses exportations et de 4% de ses importations avec le pays d'origine de ces immigrés. Nous montrons que ce rôle est d'autant plus fort que la qualité des institutions dans le pays d'origine des immigrés est faible ou que le bien échangé est complexe. Cela accrédite l'idée que ces immigrés possèdent une information spécifique sur leur pays d'origine qui permet de réduire un certain nombre des déficits d'information normalement existants.

Enfin, le dernier chapitre de cette thèse complète l'analyse en étudiant la sensibilité de chacun des exercices statistiques précédents au choix d'un système d'unités spatiales particulier. Nous étudions l'existence dans chacun des trois exercices précédents - mesure de la concentration spatiale, mesure de l'ampleur des externalités d'agglomération et mesures des déterminants spatiaux du commerce - d'une sensibilité à un changement de la taille (ou de manière équivalente du nombre) ou bien de la forme des unités spatiales qui constituent le découpage géographique sous-jacent à l'analyse. Cet exercice trouve sa pleine justification dans le fait que, la plupart du temps, les économistes sont contraints dans le choix d'une maille géographique particulière, et ne peuvent donc pas tester la robustesse de leurs résultats selon cette dimension. Nous montrons que lorsque l'agrégation n'a pas lieu à une échelle géographique trop grande, la sensibilité au changement de taille reste modérée, et la sensibilité à la forme des unités géographiques est encore moins prononcée. Ces deux distorsions sont de toutes façons de moindre ampleur que celles induites par une mauvaise spécification de l'exercice statistique.

Mots clés: Concentration spatiale, Économies d'agglomération, Productivité des entreprises, Immigration et commerce, Problème des Unités Spatiales Modifiables

ÉCOLE DES PONTS PARISTECH

Doctoral School
ÉCONOMIE PANTHÉON-SORBONNE

Laboratory
PARIS SCHOOL OF ECONOMICS

PhD DISSERTATION

in ECONOMICS

Agglomeration and the spatial determinants of productivity and trade

Anthony BRIANT

Publicly defended on April 16, 2010

PhD Advisor: Pierre-Philippe COMBES
PhD Co-advisor: Miren LAFOURCADE

Committee:

Dominique BUREAU,	Ingénieur général des Ponts et Chaussées, Professeur chargé de cours à l'Ecole Polytechnique
Pierre-Philippe COMBES,	Directeur de recherche au CNRS, Université d'Aix-Marseille
Miren LAFOURCADE,	Professeur des Universités, Université de Paris-Sud XI
Philippe MARTIN,	Professeur des Universités, Institut d'Études Politiques de Paris (referee)
Henry OVERMAN,	Reader, Department of Geography and Environment, London School of Economics
Sébastien ROUX,	Administrateur INSEE, Centre de Recherche en Economie et Statistique de l'INSEE
William STRANGE,	RioCan Real Estate Investment Trust Professor of Real Estate and Urban Economics, Rotman School of Management, University of Toronto (referee)

Contents

Introduction

The tendency of human and economic activities to agglomerate is obvious, as proved by the existence of cities. This fact is also true for individual industries. Studying the patterns of spatial concentration (or localization), i.e. detecting areas where firms in the same industry tend to cluster, is an old subject of interest for economists, dating back at least to Alfred Marshall (1890).

However, agglomeration creates congestion, bids up the price of immobile factors of production, and potentially exacerbates competition (at least when the output market is local). Several questions are thus of interest: is spatial concentration pervasive in all industries, or limited to a few anecdotal cases? What are the advantages for firms to cluster that are able to offset extra costs due to agglomeration? How large are these advantages? How fast do they decline in space? How do they shape the spatial distribution of trade? These questions make up the background picture of this dissertation.

On the origin of agglomeration

The starting point to think about agglomeration is the *spatial impossibility theorem*, proved by Starrett (1978). It states that[7] without any heterogeneity in the underlying space, and without indivisibilities or increasing returns, any competitive equilibrium in the presence of transport costs will feature only fully autarkic locations where every good will be produced at small scale (see Ottaviano and Thisse, 2004, for a detailed discussion).

A contrario, agglomeration can be observed as soon as space is heterogeneous or any indivisibilities or increasing returns exist, under the assumption that moving costs are not null.[8] Heterogeneity in space just points to the fact that places benefit from specific endowments (natural endowments, technologies or amenities) that favor their relative specialization in one type of economic activity, as in the classical trade theory. This is obviously true at a global, world wild level, but remains valid even in a small country like France. Some places may enjoy endowments that attract specific industries as the coalfields in the north-east of France. However, space tends to be more homogeneous as the geographical scale of analysis is reduced. Hence, this kind of local comparative advantages, even when widely defined, cannot explain industrial localization as a whole at the scale of a country (see for instance Ellison and Glaeser, 1999, for the US).

In this case, alternative explanations for the agglomeration of economic activities have to be found. In the presence of any moving costs, agglomeration can be explained by the ex-

[7] I quote here Puga (2009).

[8] Combes, Mayer, and Thisse (2008b, chapter 2) provide some enlightening comments on the role of space in economic thought.

istence of increasing returns or *agglomeration economies*. Such agglomeration economies exist as soon as an individual's productivity rises when he or she is close to other individuals. Agglomeration economies may be *pure externalities*, as in the case where productivity rises from being able to learn from or imitate a neighbor. These agglomeration economies can also work entirely within the market. If a supplier and a customer get closer, they may become more productive only by eliminating some kind of transaction costs, but there is no obvious externality (see Glaeser, 2008).

These returns to scale can be internal to the firm. Krugman (1991) (and the subsequent New Economic Geography literature) shows how firms tend to cluster when transport costs fall. In this setting, firms benefit from an endogenous larger market size, because they grow larger by fully exploiting internal returns to scale (see Combes et al., 2008b, for a complete overview). On the contrary, the traditional Urban Economics posits the existence of some production externalities between firms producing with constant returns to scale (see Henderson, 1974). Thus, local increasing returns, external to the firm, are at play. Firms also benefit from a larger market size or a more specialized environment depending on the type of externalities at work. This dissertation deals with this latter type of increasing returns.

When external economies are not specific to firms within an industry, they tend to agglomerate overall economic activity and give birth to cities. Hence, they are called *urbanization economies*. On the contrary, when these economies are specific to firms within an industry, they give birth to industrial clusters and are called *localization economies*. Urbanization and localization economies can be understood as a way to reduce any moving costs for goods, workers, or ideas (see Glaeser, 2008) and thus increase individual - firm or worker - productivity. Finding evidence for the existence of such agglomeration economies and quantifying their magnitude is the most fundamental question in urban economics, since without answering them we cannot understand the existence of cities and industrial clusters. Chapter 1 provides evidence for the existence of agglomeration economies, by comparing the localization patterns of French service and manufacturing industries. Chapters 2 and 3 study the magnitude of both urbanization and localization economies on the productivity of French firms.

The sources of these agglomeration economies are also a much debated issue in the literature. Various mechanisms have been suggested to explain these increasing returns to scale in cities and industrial clusters. Duranton and Puga (2004) provide an exhaustive survey of these mechanisms under the headings: sharing, matching and learning. A larger final-good industry can sustain a wider variety of input suppliers as well as a thicker labor market allowing for gains from a narrower specialization. Matching between employers and employees are supposed to be easier and of better quality in a thicker labor market. Finally, proximity between firms may facilitate the generation, the diffusion and the accumulation of knowledge. Chapter 4 provides an indirect proof for the accumulation of local knowledge when economic agents agglomerate. More precisely, we study how the spatial concentration of immigrants across French *départements*[9] impacts on their international trade flows toward the immigrants' countries of origin.

Urbanization or localization economies are, by definition, localized, geographically limited to a small area. The spatial scope of agglomeration economies is a last important dimension in this literature, as recently surveyed by Rosenthal and Strange (2004). Due to data limitations, researchers often assume that these externalities exist between firms within a administratively-defined area. However, relying on a predefined zoning system could affect statistical inference. Chapter 5 wraps this dissertation up by considering the sensitivity of the various econometric results shown in previous chapters to the choice of a specific zoning system. In other words, we study whether and how the size and shape of spatial units impact on the extent of spatial concentration, the magnitude of agglomeration economies and the spatial determinants of trade.

Localization as an evidence of agglomeration economies

Measuring excessive spatial concentration of economic activities, or individual industries, is the first sign of the existence of agglomeration economies. Measuring spatial concentration consists in describing spatial inequalities in terms of production or employment. Economists and geographers have developed a number of indices to account for spatial inequality (see Combes et al., 2008b, chapter 10).

A first approach consists in measuring, for each industry, the deviation of the cross-regional distribution of industrial employment from that of the overall employment. Traditional indices, as the Gini locational index, rest on this methodology. However, Ellison and Glaeser (1997) emphasize that industrial *lumpiness* could corrupt these traditional measures of spatial concentration. The intuition is the following: even if plants in an industry were randomly distributed over space, the cross-regional distribution of industrial employment cannot exactly match the cross-regional distribution of overall employment, due to the limited number of plants in that specific industry. They suggest an index of *spatial* concentration comparable across industries with different *industrial* concentrations, i.e. with different numbers of plants and different plant-size distributions. Their approach fulfills at least three of the six properties for an ideal index of spatial concentration listed by Combes and Overman (2004): the index is defined with respect to a clear reference, the statistical significance for spatial concentration can be assessed, the index is comparable across different industries. However, their approach relies on a discrete zoning system. Such an index is thus potentially affected by the shape, size and relative position of discrete spatial units. These problems are labelled Modifiable Areal Unit Problems (MAUP).[10] The index suggested by Ellison and Glaeser (1997) thus fails to meet two other criteria: not being sensitive to a change in the spatial zoning system and being comparable across areas.[11]

[9] Continental France is mapped by 94 départements. See chapter 4 for further details on French administrative zoning systems.

[10] Chapter 5 describes the Modifiable Areal Unit Problem in details.

[11] Note that the Ellison and Glaeser index is also sensitive to a change in industrial classification. Combes

To circumvent these limitations, a second approach answers the MAUP by considering space as continuous. The idea, initiated by Duranton and Overman (2005), is to consider the density distribution of bilateral distances between all pairs of plants in each industry. They test whether or not the observed density distribution of bilateral distances in each industry is close to the expected density distribution when plants are randomly allocated over space. They assess the statistical significance of the deviation from randomness by building a global confidence interval around this expected distribution.[12] In order to compute bilateral distances, they use the spatial coordinates of plants. This methodology is thus very data intensive. The index suggested by Duranton and Overman (2005) meets all the useful properties listed by Combes and Overman (2004), except that it remains sensitive to a change in industrial classification.

Chapter 1 builds on these two approaches and introduces a new methodology to test for localization in a continuous space. We first show that the methodology suggested by Duranton and Overman (2005) is implicitly sensitive to the industrial structure, i.e. the number and size distribution of plants within an industry. Hence, their index is not fully comparable across industries. We suggest an alternative approach that also relies on the density distribution of bilateral distances between plants within the same industry, but we build our test for localization by relying on a measure of divergence between density distributions. The idea is to estimate the divergence between the density distribution of distances within an industry and the overall density distribution of distances. This measure of divergence is not *a priori* comparable across industries, due to the small number of plants in each industry. Building on insights from the Ellison and Glaeser (1997)'s approach, we compute an index of localization comparable across industries. Our approach, relying on distances instead of employment, allows us to provide results on the spatial scope at which localization occurs. More precisely, we quantify whether the divergence between the two density distributions occurs mainly at short distances (before 4 km), at medium distances (between 4 and 40 km), or at rather long distances (between 40 and 140 km). We can thus sort out industries by the specific distance at which they appear localized. Our prior is that this distance depends on the type of agglomeration economies motivating their clustering scheme. As an example, industries where face-to-face contacts and technological spillovers are the main drivers of location choices should be localized at very short distances (at most a few kilometers). On the contrary, when labor market pooling or input sharing motives are at work, localization patterns could occur at larger distances, but still within labor market areas.

We then apply this new methodology on French data for business-oriented services and manufacturing industries. A second interest of this chapter is to systematically compare the localization patterns of service and manufacturing industries. We first find that service industries diverge more often from randomness than manufacturing industries. Second, when we turn to our per-distance analysis, we show that the majority of diverging service

and Overman (2004) suggest that an ideal index should not be sensitive to such a change.

[12]See appendix 1.7 of chapter 1 for a more technical presentation and discussion about the construction of the global confidence interval defined by Duranton and Overman (2005).

industries are localized at very short distances (before 4 km). This is consistent with the fact that services are mainly located in the heart of a few big cities.

Quantifying the magnitude of agglomeration economies

Excessive localization provides a first evidence for the existence of agglomeration economies. But how large are the benefits from agglomeration for firms? The question is to quantify the magnitude of agglomeration economies. Puga (2009) and Strange (2009) distinguish three different approaches in the literature.

The first one consists in comparing employment growth across cities or industrial clusters. The intuition is simple: if firms are more productive in cities or clusters, employment in these locations should grow more rapidly. Seminal papers by Glaeser, Kallal, Scheinkman, and Schleifer (1992) and Henderson, Kuncoro, and Turner (1995) link the long-run growth of sectoral employment in American Metropolitan Statistical Areas (MSA) to local sectoral specialization (*localization economies*) or local industrial diversity (*urbanization externalities*). Glaeser et al. (1992) conclude that urbanization externalities are prevalent and that local industrial diversity matters for sectoral employment growth. On the contrary, taking into account dynamic effects, Henderson et al. (1995) show that specialization effects prevail. For France, Combes (2000) and Combes, Magnac, and Robin (2004) study the long-run growth of employment at a smaller geographical scale, the employment area ('zone d'emploi'),[13] and disentangle the growth of employment in existing plants (the intensive margin) from the birth of new plants (the extensive margin). They show that existing plants grow more rapidly in areas with a large number of plants of various sizes, whereas new plants are mostly located in areas with a small number of plants of various sizes. These papers also underline the dynamic effect of agglomeration economies. Combes et al. (2004) show that static externalities are prevalent in France whereas Henderson (1997) finds that lagged effects still impact on sectoral growth after six to seven years. Those early studies rely on the strong assumptions that an increase in productivity due to agglomeration economies induces employment growth. This causality is not as straightforward as these studies may claim. For example, it is possible for an increase in productivity to lead to a drop in regional employment (see Combes et al., 2004, for more details). In this case, employment growth regressions are not well-suited to estimate the magnitude of agglomeration economies.

The second approach consists in comparing input returns - wages and land rents - across cities or industrial clusters. If firms are more productive in cities or clusters, they are ready to attract workers with higher wages and to pay higher land rents. Strange (2009) provides a selective literature review about the urban wage premium, i.e. the impact of agglomeration economies (especially employment density) on wages. A seminal contribution is

[13]Employment areas are spatial units underpinned by clear economic foundations, being defined by the French National Institute of Statistics and Economics (INSEE) so as to minimize daily cross-boundary commuting, or equivalently to maximize the coincidence between residential and working areas.

Glaeser and Maré (2001) who find that workers in cities larger than 500,000 have wages that are 33% higher than workers in rural areas. The urban wage premium shrinks to 5 to 11% when these authors carefully control for the unobserved heterogeneity across workers. Indeed, data on workers have the great advantage that they provide, next to wages, a bunch of individual characteristics[14] that allow researchers to deal with a number of important pitfalls. For instance, on a very rich French dataset, Combes, Duranton, and Gobillon (2008a) show that almost half of spatial disparities in wages across areas is explained by the sorting of workers according to their qualification. Cities attract more able workers. This sorting mechanism corrupts any naïve estimate of the urban wage premium, what Combes, Duranton, Gobillon, and Roux (forthcoming) name the "endogenous quality of labor" bias. Combes et al. (2008a) develop a complex two-way procedure to purge wages from any individual observable and unobservable determinants of wages, by using individual fixed effects. Remaining spatial disparities in the *net wage* are then explained by proxies for urbanization and localization economies. They find that urbanization proxies (more precisely the density of total employment) are strongly related to wages, while localization proxies (namely an index for local specialization) are statistically significant but of a weak economic effect. Combes et al. (forthcoming) extend their results by considering another source of bias, the "endogenous quantity of labour". This source of errors is due to simultaneity between wages (or more broadly productivity) and density. Indeed, high productivity locations could attract more people, and then being denser. In this case, causality runs from productivity to agglomeration. They use credible geological and historical instruments explaining disparities in density across locations, but unrelated to labour productivity. Contrary to the "endogeneous quality of labor" bias, the simultaneity bias is of small magnitude. Rosenthal and Strange (2008) also develop an instrumental variable approach to deal with the simultaneity problem on US wage data, and also conclude to a small bias. A word of caution is needed when using wages to estimate the magnitude of agglomeration economies. Indeed, according to Rosen (1979) and Roback (1982), in a context of spatial equilibrium, wages are endogenously determined by the migration of workers. Broadly speaking, if firms benefit from production externalities in cities, they will push wages up to attract workers. However, if workers enjoy some valuable consumption amenities in cities, they will be ready to accept lower wages. Wages then results from the interplay between production externalities and consumption amenities (see Glaeser, 2008, for an enlightening discussion of the spatial equilibrium concept).

The most direct way to quantify agglomeration economies is to track their impact on firm productivity. First attempts in this direction relate aggregate local measure of productivity to the size of the local market (Moomaw, 1981; Nakamura, 1985; Henderson, 1986; Ciccone and Hall, 1996; Ciccone, 2002). However, these macro-level studies are all plagued by the problems highlighted in the previous paragraph, especially the sorting of economic agents according to observable and unobservable determinants of their produc-

[14]Papers considering the impact of agglomeration economies on land rents are scarcer, mainly due to the fact that characteristics on commercial housing is more difficult to obtain.

tivity. This is the reason why researchers take benefit from the recent availability of firm level datasets to track for direct evidence of agglomeration economies on their individual productivity, controlling for a bunch of individual determinants of productivity unrelated to location. Henderson (2003) is the first to introduce in a plant-level production function some proxies for agglomeration economies. His dataset consists in a non-exhaustive panel of plants in the US, observed every five years in the machinery and high-tech sectors, with pieces of information on value-added, capital and employment for each plant. He further introduces individual fixed effects in his regression in order to capture any unobserved firm characteristics. He finds evidence for localization economies as the number of other own industry plants affects positively firm productivity in high-tech sectors but not in machinery ones. He also finds that single-plant firms benefit from and generate more external benefits than multi-plant firms.

In chapter 2, we quantify the magnitude of agglomeration economies on French firm productivity using detailed data from the tax administration. Following Combes et al. (forthcoming), we employ a two-step procedure. In the first step, we estimate a Cobb-Douglas production function whose residual is an individual productivity. We then explain disparities in average firm productivity across industrial clusters by measures of urbanization and localization economies. In the first step, the large array of individual controls provided in the tax administration data files is used to purge individual productivity from a number of its determinants unrelated to agglomeration economies, but whose omission could bias our estimates. Especially, we control for the quality of the labor force in each plant. It prevents our estimates from being plagued by errors due to the spatial sorting of workers according to their qualification. We also control for any unobservable sector-specific determinants of productivity by introducing sector-specific dummies. Indeed, some high-tech, high-productivity sectors tend to locate in larger and denser markets. This second type of spatial sorting would also make firms more productive in these areas, even if any externality were absent. In the second step, we show that firms located in the densest clusters (i.e. in the 9^{th} decile of the employment density distribution) are, on average, 8% more productive than firms located in the least dense areas (i.e. the 1^{st} decile of the employment density distribution). This effect is sizeable when compared to the 2.2% annual average productivity growth registered by French firms over 1993-1999. Not only does local density matter for firms, but also does a good access to surrounding markets, captured by market potentials. However, we only find a small, negative effect of diversity on productivity, once density is accounted for. Taken together, these results suggest that urbanization effect are of primary importance for the productivity of firms. Regarding localization economies, we show that firms located in areas of the 9^{th} decile for specialization are, on average, 5% more productive than firms located in areas of the first decile for specialization. The impact of specialization is thus less marked than the impact of density but remains important. We also find a positive and significant correlation between the quality of the labor force in a cluster and firm productivity, but this variable does not add to the explanatory power of the model. Hence, we cannot deny the existence of human capital externalities, but once controlled for the quality of the labor input in each plant, this variable does not impact on productivity, beyond the effect of density and specialization.

Recent theoretical developments (see Melitz and Ottaviano, 2008) suggest that firms can be sorted over space according to productivity, even within the same industry. This sorting effect is due, for instance, to a tougher competition in denser markets that forces the least productive firms to exit. In order to partially capture such an effect, we introduce in our second-step regression some regional dummies that control for the average firm productivity at the regional level. In this setup, the estimation relies on the comparison of average firm productivity across clusters of the same region. Even with this inclusion, our results remain robust.

In the second step of the previous procedure, we explain disparities in *average* firm productivity by agglomeration proxies. Hence, in chapter 2, we consider that agglomeration economies impact on all firms in the same way *on average*. However, even in a narrowly-defined industrial cluster, producers are heterogeneous, and the average firm productivity provides only a partial information about the whole distribution of productivities. The main goal of chapter 3 is to question that implicit assumption. We do not only consider that agglomeration economies can impact upon *the average firm productivity* but can also induce some more complex shape shifts in *firm productivity distribution* from one cluster to the other. We thus claim that heterogeneity among producers can not be disregarded in order to fully understand the impact of agglomeration economies on individual outcomes. To this aim, we use a quantile regression approach that allows us to parsimoniously quantify the impact of both urbanization and localization economies at different points in the firm productivity distribution. The semi-parametric technique of quantile regressions, introduced by Koenker and Bassett (1978) can be used to characterize the entire conditional distribution of a dependent variable given a set of regressors. Coefficient estimates at distinct quantiles may be interpreted as differences in the response to the changes in the regressors at various points in the conditional distribution of the dependent variable. Namely, in the problem under scrutiny, we can assess the impact of both urbanization and localization economies on firm productivity at different quantiles of the conditional productivity distribution. Two important results stand out from our analysis. Firms are not only more productive in denser areas, but the increase in productivity induced by agglomeration economies is stronger for the most productive firms. Firms are also more productive in more specialized areas, but localization economies, contrary to urbanization economies, do not benefit more the most productive firms. These two results question the theoretical literature on agglomeration economies. Indeed, few papers in the literature on agglomeration economies consider vertical heterogeneity across producers. Combes, Duranton, Gobillon, Puga, and Roux (2009) stands as one of the exceptions. They suggest that workers are more productive when they work for more efficient firm and that this effect is enhanced by interactions with other workers. In other words, initial heterogeneity across producers is magnified by agglomeration economies through the interplay of labor productivity. There is however no doubt that the link between agglomeration economies and vertical heterogeneity deserves further research, especially to understand the differentiated results between urbanization and localization economies put forward in this chapter.

On the local stock of knowledge

Chapter 4 deals with international trade issues. However, two important points of connexion with previous chapters can be highlighted. First, one of the most crucial aspect of agglomeration economies is about the importance of local knowledge for growth (see Lucas, 1988, 2001). To put it in a nutshell, the stock of local knowledge is a local public good that makes firms more productive. Building on the same type of intuition, chapter 4 studies whether the presence in a specific area of a large stock of immigrants impacts on the volume of international trade with their country of origin. Indeed, Rauch (2001) (among others) suggests that immigrants have specific knowledge about their country of origin that they can share locally. This stock of knowledge can help reducing the informational gap between buyers and sellers in their hosting region and country of origin, hence promoting bilateral trade opportunities. Indeed, despite the widespread availability of modern communication technologies, information costs still play a crucial role in shaping world trade patterns, as recently surveyed by Anderson and VanWincoop (2004).

Rauch (2001) underlines two other channels through which immigrants' ties to their home country may promote trade. Immigrant networks may provide contract enforcement through sanctions and exclusions, which substitutes for weak institutional rules and reduces trade costs. Immigrants can also bring their taste for homeland products, which should make their trade-creating impact even more salient on imports. In other words, transnational networks are a substitute for proximity in this case. Hence, this chapter focuses on network economies rather than agglomeration economies *per se*, but the bridges are numerous between the two approaches.

A second point of convergence between this chapter and the remaining part of the dissertation concerns the simultaneity problem highlighted in the previous section. In this chapter, the question at hand concerns the simultaneity between trade flows and immigrants' location choices. The trade-promoting effect of immigration is now well documented at the national level (see Wagner, Head, and Ries, 2002, for an extensive review). For instance, Gould (1994), Head and Ries (1998) and Girma and Yu (2002) find a significant trade-creating impact of immigrants settled in the United States, Canada, and the United Kingdom respectively. However, at the national level, there is a huge presumption that the correlation between trade and immigration could be spurious. The correlation between trade and immigration might arise from omitted common determinants such as colonial ties, language or cultural proximity, or reverse causality if immigrants prefer to settle in countries that have good trade relationships with their home country. To tackle this source of bias, we study the relationship between trade and immigration at a sub-national level, and control for all country-specific determinants of trade by fixed effects. The inclusion of country fixed effects allows controlling for the common determinants of trade and immigration at the national level. At the same time, cross-sectional variability in trade and immigration at the French *département* level provides sufficient information to identify the pro-trade effect of immigrants. We further resort to an instrumental variable approach, where lagged stocks of immigrants serve as instruments. Due to some persistence in the location choices of immigrants, we can assume that lagged stocks of immigrants partially determine current stock, but that they do not impact on trade flows anymore. We do find that immigration exerts a significant positive impact on trade: doubling the number of immigrants settled in a *département* boosts its exports to the home country by 7% and its imports by 4%.

We then evaluate the heterogeneous impact of immigrants on trade along two intertwined dimensions: the complexity of traded goods and the quality of institutions in the partner country. The fact that immigrants matter more for differentiated or complex goods can be taken as a support for the information-cost-saving channel of transnational networks, as suggested by Rauch and Trindade (2002). These authors find that South-Asian country pairs with a higher proportion of Chinese immigrants trade more with each other. They show that the trade-creating effect of Chinese networks is larger for differentiated goods than for homogeneous or reference price goods. Besides, Anderson and Marcouiller (2002) and Berkowitz, Moenius, and Pistor (2006) show that the quality of institutions impacts drastically on the volume of bilateral trade. Berkowitz et al. (2006) further point out that the quality of institutions matters more for complex commodities, which exhibit characteristics difficult to fully specify in a contract. This is the reason why good institutions may reduce transaction costs when contracts are more incomplete. Hence, through sanctions and exclusions, transnational networks could be a substitute for weak institutions, especially in the trade of complex products. Building on these insights, we disentangle the pro-trade impact of immigrants across both the quality of institutions in the partner country and the complexity of traded goods. In this respect, we emphasize two main results. First, immigrants especially matter for the imports of complex goods, regardless of the quality of institutions in the home country. Turning to the imports of simple products, immigrants matter only when the quality of institutions at home is weak. Second, the trends are less marked for exports. The pro-trade impact of immigrants on exports is positive only when they come from countries with weak institutions, regardless of the complexity of products.

The tyranny of geography: The Modifiable Areal Unit Problem

The last chapter of this dissertation makes a methodological contribution. It focuses on the Modifiable Areal Unit Problem. We pointed earlier the fact that the Ellison-Glaeser index is defined in a discrete space and thus is sensitive to the choice of a specific spatial zoning system. For instance, imagine that firms cluster in a specific area, but that an administrative boundary of the underlying zoning system crosses this area. In this case, the industry will not appear more localized according to the Ellison-Glaeser index than the same industry located in two spatial units at both ends of the country. Avoiding this kind of border effects is one of the main motivation of the distance-based approach adopted by Duranton and Overman (2005). Regarding the magnitude of agglomeration economies, some recent attempts also try to avoid this issue by considering space as continuous (see Rosenthal and Strange, 2003, 2008).

More generally, most empirical work in economic geography relies on scattered geocoded data that are aggregated into discrete spatial units, such as cities or regions. The aggregation of spatial dots into boxes of different size and shape is not benign regarding statistical inference. The sensitivity of statistical results to the choice of a particular zoning system is known as the Modifiable Areal Unit Problem (hereafter MAUP). In chapter 5, we investigate whether changes in either the size (equivalently, the number) of spatial units, or their shape (equivalently, the drawing of their boundaries) alter any of the estimates computed in previous chapters. Then, we address the important question of whether distortions due to the MAUP are large compared to those resulting from specification changes.

This exercise is all the more important because most empirical work in regional and urban economics relies on a predefined zoning system. For instance, much work has tried to check empirically whether agglomeration enhances economic performance at the scale of countries, European regions, U.S. states or even smaller spatial units such as U.S. counties or French employment areas. The magnitude of the estimates differs between papers, but we do not know whether this reflects zoning systems or real differences in the extent of knowledge spillovers, intermediate input linkages, and labor-pooling effects on firm productivity.

On the contrary, in chapter 5, we start by evaluating the degree of spatial concentration under three types of French zoning systems (administrative, grid and partly random spatial units) and by comparing the differences between concentration measures (Gini versus Ellison and Glaeser) with those between zoning systems. We then turn to regression analysis as not only is the measure of any spatial phenomenon likely to be sensitive to the MAUP, but also its correlation with other variables. We estimate the impact of employment density on labor productivity and compare the magnitude of agglomeration economies across zoning systems and econometric specifications. Finally, we run gravity regressions. We study how changes in the size and shape of spatial units affect the elasticities of trade flows within France with respect to both distance- and information-related trade costs.

All of these empirical exercises suggest that, when spatial units remain small, changing their size only slightly alters economic geography estimates, and changing their shape matters even less. Both distortions are secondary compared to specification issues. More caution should be warranted with zoning systems involving large units, however. The MAUP is obviously less pervasive when data variability is preserved from one scale to another. When moving from dots to boxes, specific attention should be devoted to the following key points: 1- the size of boxes in comparison with the original dots, 2- the way data are aggregated, i.e. averaging or summation, 3- the degree of spatial autocorrelation in the data. The MAUP is less jeopardizing when data are spatially-autocorrelated and averaged, as is the case in wage regressions. By way of contrast, the MAUP is more challenging when variables in a regression are not computed under the same aggregation process. In gravity regressions for instance, moving from one scale to another requires a summation of trade flows on the left-hand side, whereas distance is averaged on the right-hand side.

Finally, when zoning systems are specifically designed to address local questions, as is the case for French employment areas, we definitely argue that they should be used. Those who are left with other administrative units should not worry too much however, as long as the aggregation scale is not too large. We therefore urge researchers to pay attention in priority to choosing the relevant specification for the question they want to tackle.

CHAPTER 1

Location patterns of services in France: A distance-based approach[1]

1.1 Introduction

In recent years, an increasing number of empirical studies has been devoted to the description of industrial location patterns in various countries, across different time periods.[2]

Two main approaches have been followed so far. A first approach consists in measuring the deviation of the cross-regional distribution of employment in each industry from the distribution of the overall employment. Ellison and Glaeser (1997) (hereafter EG) emphasize that industrial *lumpiness* could corrupt traditional measures of spatial concentration, as Gini locational coefficient. They propose an index of *spatial* concentration comparable across industries with different *industrial* concentrations, i.e. with different numbers of plants and different plant size distributions. However, their approach relies on a discrete space.[3] It is now well understood that such an index is affected by the underlying spatial zoning system, i.e. the shape, size and relative position of spatial units. These problems are labelled Modifiable Areal Unit Problems (hereafter MAUP) in the regional science literature (see chapter 5 for more details).

A second approach answers the MAUP by considering space as continuous. The idea initiated by Duranton and Overman (2005) (hereafter DO) is to consider the density distribution of bilateral distances between all pairs of plants in each industry.[4] They test whether or not the observed density distribution of bilateral distances in each industry is close to the expected density distribution when plants are randomly allocated over space. They assess the statistical significance of the deviation from randomness by building a global confidence interval around this expected distribution.[5]

[1]This paper is a joint work with Muriel Barlet (DREES-INSEE) and Laure Crusson (DARES-INSEE).

[2]Several surveys focus on this question in the last volume of the Handbook of Urban and Regional Economics: for North America (Holmes and Stevens, 2004), for Europe (Combes and Overman, 2004) and for Japan and China (Fujita, Mori, Henderson, and Kanemoto, 2004).

[3]Along this route, other papers can be mentioned: Maurel and Sédillot (1999), Devereux, Griffith, and Simpson (2004), and Guillain and Le Gallo (2007) to name a few.

[4]The number of papers considering space as continuous is relatively scarce, but increasing. Duranton and Overman (2008) provide further results for the UK, and Klier and McMillen (2008) provide an application to the US auto supplier industry. Marcon and Puech (2003), Marcon and Puech (2007) and Arbia, Espa, and Quah (2008) propose another approach in continuous space based on point-pattern analysis (Diggle, 2003). Arbia, Espa, Giuliani, and Mazzitelli (2009) study both the temporal and spatial scopes of agglomeration.

[5]See appendix 1.7 for a more technical presentation and discussion about the construction of the global

In this paper, we continue along this second route. We show that the test suggested by Duranton and Overman (2005) suffers from a systematic upward bias and, more importantly, that this bias increases with the number of plants in the industry. The test is thus implicitly sensitive to industrial concentration. Hence, it is less attractive for comparisons across industries. The first contribution of our paper is to propose a new test for localization in continuous space, which is fully comparable across industries. We apply this new methodology to compare the location patterns of French business services and manufacturing industries, for which industrial concentrations drastically differ.

We develop a two-step procedure to characterize location patterns in continuous space. We first pick out industries whose density distribution of bilateral distances departs significantly from randomness, using a test based on a measure of divergence in the space of density distributions.[6] We detect the so-labelled *diverging industries* for which location choices are not random but driven by agglomeration or dispersion forces. Among these diverging industries, we then disentangle *localized industries*[7] from *dispersed ones*. Broadly speaking, an industry is said to be localized if its plants are closer to one another than randomly allocated plants of an hypothetical industry with the same industrial concentration. This definition of localization implicitly depends on a specific threshold distance d. We can thus sort out industries by the specific distance at which they appear localized. Our prior is that this distance depends on the type of Marshallian agglomeration forces motivating their clustering scheme. As an example, industries where face-to-face contacts and technological spillovers are the main drivers of location choices should be localized at very short distances (at most a few kilometers). On the contrary, when labor market pooling or input sharing motives are at play, localization patterns could occur at larger distances. In our view, such a sorting of industries provides a first pass in the identification of the specific mechanisms underlying agglomeration economies. Finally, for a large enough distance d, the industry is said to be dispersed, rather than localized at long distance. In this case, dispersion forces overcome agglomeration ones.

The second contribution of this paper is to focus on the location patterns of business-oriented service industries. Services play an increasing role in terms of employment and value-added in modern economies. However, following a long tradition dating back to Marshall (1890) at least, the previous literature has almost exclusively studied the location patterns of manufacturing industries. The main reason for this bias is the belief that output of service industries is highly non-tradable and that these industries are stuck in their location choices to their customers. What remains certainly true for personal health care services or retail activities is less obvious for business-oriented services. As suggested by Kolko (1999), the improvement in communication technologies can make the delivery of

confidence interval.

[6]Note that Mori, Nishikimi, and Smith (2005) also use a measure of divergence to test for significant spatial concentration in a discrete space. They compare the cross-regional distribution of sectoral employment to the cross-regional distribution of surface areas.

[7]We use in this paper the concept of *localization* in a continuous-space setting, and the concept of *spatial concentration* in a discrete-space framework.

digitalized service output easier. Freed from their physical proximity to customers, these industries can fully benefit from localization economies, as the sharing of a highly skilled workforce or across-the-street networking (Arzaghi and Henderson, 2008). These localization economies are certainly all the more important for these industries because other determinants of location choices, as local endowments or local indivisible facilities, are less prevalent. It is the reason why we focus our study on business-oriented services.

The systematic comparison between the patterns of localization and dispersion of business-oriented services (hereafter services) and manufacturing industries provides some striking results. We first find that service industries diverge more often from randomness than manufacturing industries. Second, when we turn to our per-distance analysis, we show that the majority of diverging service industries are localized at very short distances (before 4 km). This is consistent with the fact that services are mainly located in the heart of a few big cities[8] and strongly benefit from very localized technological spillovers. This result has not been emphasized in the previous literature, except for anecdotal highly-localized industries as advertising activities in Manhattan (Arzaghi and Henderson, 2008) or casino hotels in Las Vegas (Holmes and Stevens, 2004). Third, we show that the localization patterns of services are mainly driven by the specific location choices of the largest plants. For a majority of service industries, these largest plants are more localized than the overall plants in the industry. This is not true for manufacturing. This result partially confirms findings by Holmes and Stevens (2002). Fourth, when turning to dynamics, tendencies put forward for manufacturing industries (Dumais, Ellison, and Glaeser, 2002) are far much stronger for services. New plants tend to disperse service industries by locating outside of existing clusters. On the contrary, the dispersion of exiting plants tend to reinforce the localization patterns of most service industries.

The few previous studies concerning services all rely on a discrete-space approach using the EG index (see for instance Kolko, 1999, forthcoming; Holmes and Stevens, 2004). However, services are most often located in the heart of big cities, where economic activity is densely agglomerated and diversified. Mori et al. (2005) argue that the EG index undervalues the concentration of industries mostly located in highly agglomerated and diversified areas.[9] To support this argument, table 1.1 provides moments for the distribution of the EG index (computed at the employment-area[10] level) for both French manufacturing

[8]Kolko (1999) shows that the share of business services employment is relatively larger in the US big cities than the share of manufacturing industries, and that this share has increased more rapidly in big cities between 1977 and 1995.

[9]In appendix A of their paper, Mori et al. (2005) develop an example to illustrate this point. The intuition is simple. If regions are not to small, it will be more difficult for an industry whose employment is mostly located in dense regions to deviate from the overall employment distribution. In other words, industries mostly located in rural areas will register a higher value for the EG index. For instance, it is the reason why agriculture almost always appear concentrated according to the EG index.

[10]The French continental territory is partitioned into 341 employment areas. These spatial units are underpinned by clear economic foundations, being defined by the French National Institute of Statistics and Economics (INSEE) so as to minimize daily cross-boundary commuting, or equivalently to maximize the coincidence between residential and working areas.

and service industries. Contrary to the results uncovered with our continuous approach, services appear much less concentrated than manufacturing industries with the EG index on average.[11] Our distance-based method does not suffer from the drawback put forward by Mori et al. (2005) and is thus better suited to extract meaningful features concerning the location patterns of services.

Table 1.1 – Distribution of the EG index for service and manufacturing industries

	Mean	Median	Variance	Min	Max
Manufacturing Industries	0.040	0.019	0.003	-0.032	0.403
Service Industries	0.028	0.010	0.003	-0.189	0.238

Notes: The EG index is computed at the employment-area level using the 407 4-digit items of the French industrial classification (NAF700) that we use throughout this paper. See section 1.3 for further details.

Distance-based methods are data intensive. They require precise information on the geographical coordinates of plants, so as to accurately compute bilateral distances and to study location patterns at pretty short distances. In this study, we use a geo-referenced dataset newly-developed by the French Statistical Institute (INSEE). It provides the precise geographical coordinates for plants located in cities with more than 10,000 inhabitants. When this information is missing, we consider that plants have the same geographical coordinates as the city hall of their municipality of location. Municipalities constitute the finest French spatial division. As an illustration, the French continental territory is covered by more than 36,000 municipalities. It allows us to reduce measurement errors in the computation of the remaining bilateral distances.

The rest of the paper is organized as follows. In the next section, we introduce our new test for localization in continuous space. We then present more precisely the French geo-referenced data at hand, before turning to our main results in sections 1.4 and 1.5.

1.2 Methodology

In this section, we describe the setup for our analysis. We first present a new way to test whether the location choices of plants within an industry significantly diverge from randomness. We then introduce a strategy to disentangle whether this divergence corresponds to localization or dispersion. For the sake of simplicity, we first present our methodology in the unweighted case (i.e. by only considering the number of plants in each industry), before turning to the weighted case (i.e. when both the number of plants and the employment-size distribution in the industry are taken into account).

[11]Even the median and maximum values of the EG index for services are lower than the corresponding quantities for manufacturing industries. Holmes and Stevens (2004) and Kolko (1999) also register a lower average value for the EG index (computed at the US county level) for service industries than for manufacturing industries. Kolko (1999) finds that the mean EG index for business service industries (at the SIC 4-digit level) (0.0064) is twice smaller than the mean EG index for manufacturing (0.0129).

1.2.1 Testing whether industrial location patterns significantly diverge from randomness

Consider an industry i with N_i plants. We first compute the $\frac{N_i(N_i-1)}{2}$ distances between all pairs of plants in that industry. We estimate the density distribution of bilateral distances between plants (hereafter *observed distribution*, f^{obs}) within the industry under scrutiny. Our goal is to evaluate how far this distribution is from a reference distribution representing randomness. For this purpose, we rely on *a measure of divergence* in the space of density distributions.

Choice of a reference distribution

The first step consists in choosing the relevant reference distribution (hereafter f^{ref}), i.e. the random benchmark against which divergence is assessed. We consider the distribution of bilateral distances between pairs of all *active sites*. Following Duranton and Overman (2005), we define the set of *active sites* as the whole set of locations where a plant is currently located, regardless of its industry. In other words, we compare the density distribution of bilateral distances within an industry to the overall distribution of bilateral distances.

We implicitly assume that the set of potential locations for a plant is limited to the set of currently existing locations. It leads us to exclude a lot of other alternatives. In their discrete-space approach, Mori et al. (2005) choose as a benchmark the *economic area*, defined as the whole surface area minus marshes, mountains and rivers. However, in a continuous-space framework, it would lead us to consider either an infinity of potential points or to restrict the possibilities to an arbitrary set of randomly allocated points. The set of active sites has, on the contrary, the main advantage to control for the observed agglomeration of the overall economic activity in France.

Estimating the observed and reference distributions

The second step consists in obtaining an estimation of the observed and reference density distributions. We rely on the 1-km bin density histogram. In our data for France, the maximum distance between two plants is equal to 1109 km. Then, our density histograms have 1110 bins of 1-km each from 0 to 1109 km. We thus count the share of bilateral distances within the range $[0, 1[$ km, $[1, 2[$ km and so on, until $[1109, 1110[$ km. The observed density distribution is then estimated by:

$$f^{obs}(d) = \frac{1}{N_i(N_i-1)/2} \sum_{i=1}^{N_i} \sum_{j=i+1}^{N_i} \mathbb{I}_{d \leq d(i,j) < (d+1)}, \tag{1.1}$$

where N_i is the number of plants in the industry and $d(i, j)$ the distance between plants i and j. d lies between 0 and 1109. $\mathbb{I}_{d \leq d(i,j) < (d+1)}$ is a dummy equal to 1 if the distance between plants i and j stands between d and $d + 1$ kilometers. It is worth noting that the width of the bin could have been smaller. However, errors in location prevent us from getting a more accurate measure of distances. Defining narrower bins would have increased the errors in frequency counts without really adding more information. Moreover, over the $[0 - 1110[$ km range, 1-km bins seem to be of sufficiently high precision.

A measure of divergence between density distributions

In order to assess how far the observed and reference distributions are, we use a *measure of divergence* in the space of density distributions. Our *measure of divergence* (or, simply, divergence) between f^{obs} and f^{ref} is defined as:

$$D^{obs} \equiv d(f^{obs}, f^{ref}) = \int_0^{1110} (f^{obs}(x) - f^{ref}(x))^+ dx. \tag{1.2}$$

We define the operator $(g(x))^+$ so that $(g(x))^+ = g(x)$ if $g(x) > 0$, and $(g(x))^+ = 0$ otherwise. In other words, D^{obs} is the area between the two densities f^{obs} and f^{ref}, when the former is above the latter. Note that, as long as f^{obs} and f^{ref} are density distributions, D^{obs} is also equal to half the sum of the absolute difference between the two distributions computed over the whole set of distances:

$$D^{obs} = \frac{1}{2} \int_0^{1110} |f^{obs}(x) - f^{ref}(x)| dx. \tag{1.3}$$

This measure has two valuable properties for our purpose. First, for any f^{obs} and f^{ref}, this metric has a lower bound (0) and an upper bound (1). This matters to define our *index of divergence*. Second, it is asymmetric $(d(f^{obs}, f^{ref}) \neq d(f^{ref}, f^{obs}))$ when computed on a subset of distances, let say $]0, d]$. This last property is necessary below to disentangle localization from dispersion.

Testing for the significance of divergence

The last step consists in testing whether the observed divergence D^{obs} is statistically significant. We compare it with the value of divergence the industry would register if its plants were randomly allocated across all active sites. We thus compute the divergence between the reference distribution and the distribution of bilateral distances for such an hypothetical industry.[12] We note $D^{sim} \equiv d(f^{sim}, f^{ref})$ the measure of divergence between a simulated density distribution (f^{sim}) and the reference distribution.

We repeat the random allocation procedure 1000 times and successively compute the values of divergence for the 1000 simulated distributions ($(f^{sim})_{0,..1000}$). Finally, for each industry, we are able to define an *empirical distribution for the measure of divergence under the null hypothesis of random allocation of plants across active sites*.

We then define the 95% percentile in this distribution. If the divergence between the reference and observed distributions (D^{obs}) is above this cut-off point, then we conclude that the industry significantly diverges from the reference distribution. This procedure provides a simple test for significance.

[12] In the simulation procedure, plants also keep their employment size. This is useful when one turns to an employment-weighted test for divergence. See part 1.2.4. Note also that this procedure of simulation is the same as the one used by Duranton and Overman (2005).

1.2.2 A distance-dependent test for localization

The test we propose in the previous section is a test for significant divergence on the whole set of distances, $[0, 1110[$ km. It is not a test for localization. In a continuous-space framework, localization is a concept that can only be defined with respect to a specific threshold distance d. Broadly speaking, an industry is said to be localized if its plants are relatively more numerous than overall active sites within this threshold distance. As a consequence, f^{obs} is above the reference distribution (f^{ref}) for some distances smaller than this given threshold. This implies that any positive value for the divergence D^{obs} is mainly driven by a positive gap between the observed and reference distributions between 0 and d. On the contrary, dispersion occurs when the divergence is mainly driven by a positive gap beyond d. We thus introduce in this paragraph a distance-dependent test for divergence that disentangles localization from dispersion.

Let us define:

$$D^{obs}(d) = \int_0^d (f^{obs}(x) - f^{ref}(x))^+. \tag{1.4}$$

$D^{obs}(d)$ is the area between the two densities f^{obs} and f^{ref}, when the former is above the latter on the range of distances $[0, d]$ km. As previously, we can carry out a 5%-level significance test for the divergence on the $[0, d]$ km range.

An industry is said to be localized at distance d if its divergence on $[0-d]$ km is greater than the 95th percentile of the distribution of divergences on $[0-d]$ km across the 1000 simulations. More generally, for any given threshold distance d, an industry is said to be localized at distance d if the null hypothesis of the corresponding distance-dependent test is rejected. On the contrary, an industry is said to be dispersed beyond d if the null hypothesis of the distance-dependent test is accepted (i.e. the observed distribution on the $[0, d]$ km range is below the 95th percentile).

Our methodology is easy to implement and allow to test for localization at many different threshold distances d. In the empirical part, we are able to propose a sorting of industries according to the specific distance at which they appear localized.

1.2.3 An index of divergence

In some applications, one needs a scalar to quantitatively assess how much an industry diverge from randomness.[13] We build in this section such a scalar. The raw measure of divergence D^{obs} is not comparable across industries with different number of plants, only the two previous significance tests are. In this paragraph, we build upon this measure an *index of divergence* fully comparable across industries. As made clear below, our approach is close to the one proposed by Ellison and Glaeser (1997).

We define our *index of divergence* δ as:

$$\delta = \frac{D^{obs} - D^{rand}}{D^{max} - D^{rand}}, \tag{1.5}$$

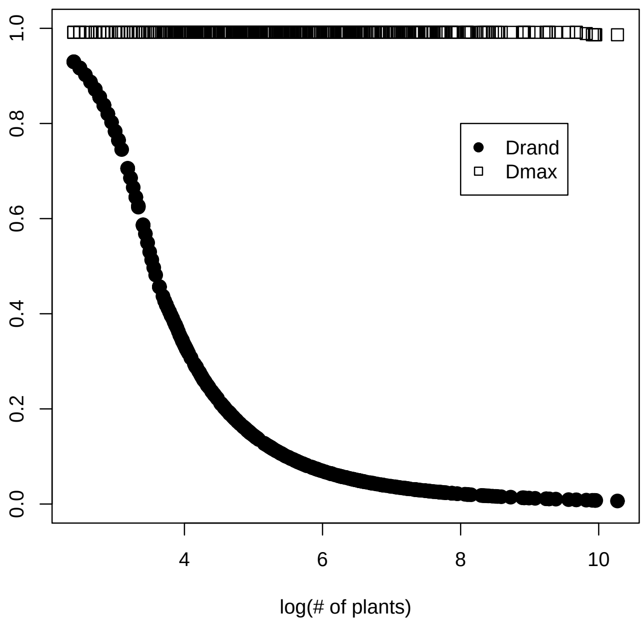
[13]See for instance Ellison, Glaeser, and Kerr (2010) for a recent use of the localization index proposed by Duranton and Overman (2005).

where D^{rand} is the average divergence across the 1000 simulations (hereafter the *average divergence*, $D^{rand} \equiv \overline{D(f^{sim})}$) and D^{max} is the achievable maximum value for D^{obs}.

Expected divergence under the null hypothesis of randomness

By definition, the divergence between f^{ref} and any simulated distribution (f^{sim}) is positive. The average divergence D^{rand} is thus strictly positive. Intuitively, the smaller the number of plants in the industry, the larger the average divergence. Indeed, when the number of plants in an industry is small, it is more difficult for the observed density distribution of bilateral distances to match the reference distribution. Figure 1.1 shows that the average divergence decreases monotonically with the number of plants in the industry. By construction, under the null hypothesis (H_0) of randomness, D^{rand} is the expected value of D^{obs}, $\mathbb{E}(D^{obs}) = \overline{D(f^{sim})} = D^{rand}$. As a consequence, the raw measure of

Figure 1.1 – D^{rand} and D^{max} as a function of the industry size



divergence D^{obs} is not a relevant measure for cross-industry comparisons. Without any agglomeration forces at play, an industry with a small number of plants will register a large value for D^{obs}. A meaningful measure of divergence is thus a relative one.

Defining a relative measure of divergence

The quantity $D^{obs} - D^{rand}$ constitutes a first pass. However, such a measure still depends on the number of plants in the industry. The expected range of variation of $D^{obs} - D^{rand}$ is $[0, D^{max} - D^{rand}]$, where D^{max} is the achievable maximum value for D^{obs}. We prove below that 1 is a good approximation for D^{max} (whatever the number of plants in the industry). Consequently, the expected range of variation for $D^{obs} - D^{rand}$ depends also on the number of plants (through D^{rand}). For instance, small industries with a large D^{rand} can not achieve a large value for the quantity $D^{obs} - D^{rand}$. A way to define a relative (i.e. unrelated to the number of plants) index of divergence is to divide this quantity by the whole range of admissible variation for D^{obs}, $D^{max} - D^{rand}$, as in equation 1.5.

Equation 1.5 can be rewritten as follows:

$$D^{obs} = D^{rand} + \delta(D^{max} - D^{rand}) = (1-\delta)D^{rand} + \delta D^{max}. \tag{1.6}$$

D^{obs} can then be understood as the barycenter between the point D^{rand} with weight $(1-\delta)$ and the point D^{max} with weight δ. These weights (and thus, δ) are unrelated to the specific number of plants in the industry under scrutiny.

More technically, a last important step is to assess D^{max} the achievable maximum value for D^{obs}. We argue that 1 is a good approximation of D^{max}, regardless of the number of plants in the industry. We have already mentioned that 1 is an upper bound of D^{obs}. In order to obtain a lower bound for D^{max}, we compute D^{obs} for a very localized industry. For each industry i with N_i plants, we allocate the N_i plants in the smallest possible circle. The value of D^{obs} for that particular spatial configuration is very close to 1. We present that value for the French industries under scrutiny on figure 1.1. For France, 90% of the industries have a number of plants sufficiently small to be located in the same circle with a one-kilometer diameter. For the largest industries, the diameter of the circle has to be slightly increased, but the value of the empirical D^{max} remains very close to one (see figure 1.1).

1.2.4 Employment-weighted test and index

So far, we have only considered the number of plants in the industry and not their employment sizes. However, Ellison and Glaeser (1997) argue that both the number of plants and the employment-size distribution have to be taken into account to properly control for the industrial concentration.

In this part, we develop a test for divergence and an index of divergence taking into account the employment-size distribution. In this new setup, the observed distribution does not describe the distribution of bilateral distances between pairs of plants, but between pairs of employees in different plants (hereafter *employment-weighted* observed distribution). In this case, large plants mainly drive the shape of the observed density distribution. Indeed, consider two large plants, the bilateral distance between these two plants has a frequency count equal to the cross-product of their employment sizes. When we turn to the density, the weight of a given distance d between two plants equal to $\frac{w_k w_l}{\sum_{k\in i}\sum_{l\in i, l>k} w_k w_l}$ where w_l (resp. w_k) is the employment in plant l (resp. k) in industry i. Except for this change, the setup remains the same. We build as previously a measure of divergence between the employment-weighted observed distribution and the reference distribution.

In this case, the average divergence across 1000 simulations (D^{rand}) does not only take into account the number of plants in the industry, but also the employment distribution across plants within the industry. Indeed, each simulation is a random allocation across active sites of plants for an hypothetical industry with the same number of plants *and* the same distribution of employment across plants. Then, δ_W[14] is an index of divergence comparable across industries with different numbers of plants *and* different employment distributions across plants.

1.2.5 Comparisons with the previous literature

In this section, we compare our approach with two important papers in the literature on location patterns: the papers by Duranton and Overman (2005) and by Ellison and Glaeser (1997).

A comparison with the methodology suggested by Duranton and Overman (2005) The main departure between our methodology and the one suggested by Duranton and Overman (2005) concerns our significance test for divergence. We argue that our test and index do not depend on the number of plants in the industry (and/or the employment size distribution across plants in the case of the weighted index). On the contrary, we empirically prove in appendix 1.7 that the DO test for localization suffers from a systematic upward bias in small samples, and, more importantly, that this bias increases with the number of plants in the industry. This introduces in the DO approach a non-trivial dependency to the number of plants. Our test does not suffer from such a bias, as proved in appendix 1.7. This point is crucial when one turns to the comparison of location patterns for industries whose number of plants drastically differ. This is the case between French service and manufacturing industries. The former industries register an average number of plants 5-time larger than the average number of plants in the latter ones.

Let us emphasize that trying to purge any measure of localization from dependency to industrial concentration is not a purely methodological exercise, but is economically meaningful. Indeed, some authors (see for instance Ellison et al., 2010) regress the DO measure of localization on proxies for Marshallian externalities. In such an exercise, the identification relies on the cross-industry variation in the measure of localization only.[15] If that measure depends on the number of plants in the industry, the results of such exercises could be questioned. At least, this number of plants should be introduced as a covariate in the regression. However, as long as the size-dependency is not linear, one would prefer to use a measure of localization independent of any size effect.

Second, Duranton and Overman (2005) estimate the density distribution of bilateral distances in a given industry with a kernel smoothing procedure *à la* Silverman (1986). They argue that smoothing is a way to erase plant location errors in the computation of distances. They choose a bandwidth based on the Silverman (1986)'s *rule of thumb*.[16] We argue that such a method is not very well suited in the French case under scrutiny. First,

[14]In the following, we distinguish δ_U the unweighted index from δ_W the weighted index.

[15]See also chapter 10 of Combes et al. (2008b) for critics on such an empirical approach.

[16]The Silverman (1986)'s *rule of thumb* is given by $0.9min(sd, R/1.34)N^{-0.20}$ where sd is the standard deviation and R the interquartile distance in the vector of bilateral distances, and N the number of bilateral distances in the industry.

in our case, the short distances are rather well measured. As made clear below (see part 1.3), errors in locations are limited. Hence, the measurement errors between the great circle distance we compute and the real distance by the road between two plants is very small as long as these two plants are not too far away. At least, this error remains small in comparison with the size of the bandwidth induced by the Silverman (1986)'s rule of thumb. For instance, the bandwidth is greater than 20 km for 60% of our industries. For larger distances, using a smoothing approach can be justified. However, approximations introduced by such a smoothing procedure depends on the shape of the density distribution (specifically its convexity) and are difficult to control for. These are the two main reasons why we rely on an histogram approach.

An analogy with the methodology suggested by Ellison and Glaeser (1997)

To firmly confirm that our index of divergence is fully comparable across industries, let us develop an analogy with Ellison and Glaeser's (1997) approach.

These authors emphasize that even if plants were randomly distributed over space, the cross-regional distribution of sectoral employment cannot exactly match the cross-regional distribution of overall employment, due to the limited number of plants in that industry.

To write it with their notation, under the null hypothesis of random distribution, the expected value of their raw concentration index G is not equal to zero[17] but to a strictly positive value $(1-S)H$, where H is the Herfindhal index in the industry and $S = \sum_{i=1}^{R} x_i^2$ if $(x_i)_{i=1..R}$ is the vector of overall employment share across the R regions. In our setting, D^{rand} plays the same role as $(1-S)H$ in the EG approach, by taking into account the fact that even under randomness, the density distribution of bilateral distances cannot perfectly match the reference distribution.

Furthermore, in the EG approach, the maximum value of the expected measure of concentration $\mathbb{E}(G)$ is equal to $1-S$. $\mathbb{E}(G)$ is maximum for a spatial configuration where all the plants of a given industry are located in the same region. This corresponds to the case where the spillovers or natural advantages are so strong that all the plants choose to locate in the same area. In this configuration, their index of concentration (hereafter, γ) is equal to 1. In our setting, we prove that for some agglomeration forces sufficiently strong, the measure of divergence between the observed and reference distributions can be made very close to 1, the upper-bound of the metric.

Ellison and Glaeser (1997) show that:

$$\frac{\mathbb{E}(G)}{(1-S)} = (1-\gamma)H + \gamma = H + \gamma(1-H). \tag{1.7}$$

By dividing equation 1.6 by D^{max}, we get:

$$\frac{D^{obs}}{D^{max}} = \frac{D^{rand}}{D^{max}} + \delta(1 - \frac{D^{rand}}{D^{max}}). \tag{1.8}$$

[17]Recall that $G = \sum_{i=1}^{R}(s_i - x_i)^2$ where $(s_i)_{i=1..R}$ is the vector of industrial employment share across the R regions, and $(x_i)_{i=1..R}$ the vector of overall employment share across the same R regions.

The analogy is then obvious. In our set-up, the quantity $\frac{D^{rand}}{D^{max}}$ plays the same role as the Herfindhal index H for Ellison and Glaeser (1997). As already mentioned, the interpretation is the same. The Herfindhal index corrects their measure of concentration for the lumpiness of employment distribution across plants within the industry. In our setting, the quantity $\frac{D^{rand}}{D^{max}}$ also takes into account that an industry with a small number of plants can not exactly match the reference distribution under the null hypothesis of random distribution.

1.3 Data

In this section, we present the data at hand to study the location patterns of service and manufacturing industries. Results are presented in the next two sections.

Raw data: location, industry and employment

Our approach relies on an estimate of the density distribution of bilateral distances between pairs of plants in each industry. Distances between plants are computed as the great circle distance.[18] We thus need to know the accurate spatial location of plants. Our empirical analysis mainly uses two exhaustive plant-level datasets available at the French Institute for Statistical and Economic Studies (INSEE) for the year 2005.

The first dataset, called the SIRENE repository, provides three pieces of information for each plant: an identification number, its main industry of activity in the 4-digit French industrial classification (NAF 700),[19] and third, its spatial location.

For plants located in a municipality with more than 10,000 inhabitants, the spatial location consists in the geographical coordinates of the plant, in the *Lambert 93* geo-referencing system. Approximatively, 50% of the plants in our sample are located in such a municipality. Geographical coordinates are thus available for these plants. For the remaining plants, the SIRENE dataset provides the complete address with number, street, municipality and zipcode. We use a software allowing to recover the geographical coordinates of a plant using its postal address. Finally, for the 10% remaining plants we are not able to recover the accurate geographical coordinates. These plants are mainly located in rural municipalities. We choose to consider that these plants are located at the geographical coordinates of the cityhall of their municipality. In French rural municipalities, economic activity is most of the time organized around the cityhall. Moreover, the whole continental French territory is covered by 36,247 municipalities, called *communes*. These *communes* are very small spatial units, whose median surface area equals to 10.7 km^2. We can thus consider that the induced error in location is not a major source of bias in the empirical analysis.

The second dataset is built upon the Annual Social Data Declarations files (DADS files). These data are collected from all employers and self-employed in France for pension,

[18] Of course, a more accurate measure of distance would have been a real distance by road. However, such a measure is unavailable. In the case of France, whose territory is mostly connected, the great circle distance is a pretty good approximation for the real distance. See Combes and Lafourcade (2005).

[19] We use the 2003 revision of this classification.

benefits and tax purposes (see Combes et al. (forthcoming) for further details). Nevertheless, for registration simplicity, multi-plant firms are allowed to register all their employees in the same plant. For these firms, the accurate location of their employment is unknown. It is the reason why we use a modified version of this dataset, called the CLAP dataset, which is dedicated to the location of each employee to the plant he really works for. These files provide us with the exact location of employment for the year 2005.

We merge these two datasets by the plant identification number. We only keep plants with at least one paid employee in 2005. Finally, we keep manufacturing industries (corresponding to the items B to F of the 1-digit French industrial classification (NAF36)) as well as business-oriented service industries (corresponding to the items K, L and N of the 1-digit French industrial classification (NAF36)). We exclude from the dataset retail and wholesale trade, real-estate services, as well as household services[20]. This allows us to reduce the sample at hand, without losing relevant information. We assume that the location of those services is mainly determined by the location of consumers, and thus not of primary interest.

Industrial concentration for manufacturing and service industries

Finally, we get information for 518,036 plants, among which 167,652 belong to manufacturing industries and 350,384 to service industries. These plants employed 3,160,055 workers for manufacturing industries and 4,230,229 workers for service industries. Consequently, while our selection of service industries may appear restrictive, service industries under scrutiny represent a larger share of employment than overall manufacturing industries.

We build our index such that it is not sensitive to industrial concentration. Such a property is meaningful as service and manufacturing industries drastically differ by their industrial concentration. The main difference between service and manufacturing industries is their number of plants. On average, manufacturing industries contain 540 plants whereas this figure stands at 3650 for service industries. Indeed, the available industrial classification is broader for service industries. There are 311 manufacturing and 96 service industries.

The second main difference between service and manufacturing industries is the mean employment size of plants within industry. This mean size is bigger for manufacturing plants which employ 19 workers on average against only 12 for services. However, what really matters is not the average-employment per plant but the heterogeneity in plant sizes. A relevant indicator of that heterogeneity is the percentage of employees hired by the 10% largest plants. This figure stands at 55% for manufacturing industries and 58% for service industries.[21] This means that within service industries the plant size distribution is slightly more heterogeneous than within manufacturing industries.

[20]These excluded items are more or less similar to the followings: Arts, Entertainment, and Recreation, Accommodation and Food Services and Other Services in the 2002 2-digit US NAICS classification.

[21]If we consider the 5% largest plants, the percent of employees are 45% for service industries and 40% for manufacturing industries. With the 20% largest plants it is respectively 73% and 72%.

1.4 Cross-industry results

In this section, we present results on the extent of localization and dispersion across service industries in France for the year 2005. We stress the comparison between the results for the manufacturing and service industries.

1.4.1 Result #1: Uneven patterns of location are more pervasive for business services than for manufacturing industries

We compute our test for divergence and index of divergence for each industry in the 4-digit French industrial classification (407 industries) for the year 2005. Table 1.2 reports the results for both the weighted and unweighted approach. The share of diverging industries

Table 1.2 – Summary statistics

	Unweighted Index (δ_U)		Weighted Index (δ_W)	
	Manuf.	Services	Manuf.	Services
Share of significantly diverging industries...				
...at the 5% level	68%	96%	58%	82%
...at the 10% level	75%	97%	62%	82%
Employment-weighted share of significantly diverging industries...				
...at the 5% level	89%	100%	76%	87%
...at the 10% level	92%	100%	79%	87%
Moments of the index distribution				
Mean of δ	0.083	0.103	0.078	0.145
Median of δ	0.048	0.055	0.043	0.086
Variance of δ	0.014	0.015	0.014	0.024
Min of δ	-0.104	-0.048	-0.099	-0.052
Max of δ	0.578	0.876	0.609	0.785

is larger among services than among manufacturing industries for both the unweighted and weighted tests (and at both the 5% and 10% levels). With the unweighted test at the 5% level, 96% of service industries (92 industries) are significantly diverging, while this figure stands only at 68% for manufacturing industries (213 industries). As first outlined by Duranton and Overman (2005), we notice that uneven patterns of location are less prevalent for manufacturing industries than previously believed (Ellison and Glaeser, 1997; Maurel and Sédillot, 1999; Devereux et al., 2004). More importantly, a new fact comes up with our data: uneven patterns of location are more pervasive for business services than for manufacturing industries.

The number of manufacturing industries is much larger in our sample than the number of service industries (311 against 96). We thus present in the second part of table 1.2 the employment-weighted share of service (respectively manufacturing) industries that diverge from randomness. Results are qualitatively the same, the share of workers employed in diverging industries is larger for services than for manufacturing industries, in both the weighted and unweighted approach. However, the difference is less pronounced, suggesting that non-diverging manufacturing industries are rather small in terms of employment.

Another result stands out from table 1.2. The extent of divergence is also on average larger for services than for manufacturing industries. The average divergence across service industries is almost twice larger than its value across manufacturing industries for the weighted index (0.145 against 0.078).[22] This contrasts with what is found with an EG index (see table 1.1 in the introduction). The distance-based approach appears more appealing to study location patterns of business services than the EG discrete index.

Which are the most diverging service and manufacturing industries?

Table 1.3 lists service and manufacturing industries with the 10 largest value of δ_U. The 10 most diverging service industries can be easily sorted into two groups. The first group consists in a set of industries for which technological spillovers or labor market pooling can intuitively drive location choices. We find in this group: re-insurance industry (660F), administration of financial markets (671A), market research and public opinion polling (741E), and data base activities (724Z). The second set of industries consists in transport activities which rely on large infrastructures as ports and airports. In this case, natural advantage (access to the sea) or indivisible facility sharing can be advocated to explain their diverging location patterns.

Concerning manufacturing, 5 out of 10 of the most diverging industries correspond to clothing industries, a result reminiscent of the findings by Duranton and Overman (2005) for the UK.[23] The best benchmark we get for France is the study by Maurel and Sédillot (1999). Using their own index, they also point out that a majority of the most spatially concentrated manufacturing industries in France are textile or clothing industries, as well as media-related industries.[24] The results with the weighted index (δ_W) are almost the same (see table 1.15 in appendix 1.8 for a list of the 10 most diverging service and manufacturing industries according to δ_W). The rank correlation between both indices stands at 0.72, whereas the Pearson correlation equals 0.85.

Note finally that for both services and manufacturing industries, the most diverging industries register a rather small number of plants in comparison with the average number of plants in each group (540 plants on average per industry for manufacturing industries and 3650 for services). Devereux et al. (2004) and Duranton and Overman (2005) put also forward the the rather small number of plants in the most spatially concentrated manufacturing industries in the UK.

[22]This conclusion remains valid when we remove the service industry with the largest index, the reinsurance industry, from our computation of the mean and the median. This result is thus not driven by this outlier.

[23]In their table 2, 6 out of the 10 most localized industries are textile industries.

[24]Maurel and Sédillot (1999) also outline that extractive industries are highly spatially concentrated. Location choices in these industries are influenced by the availability of raw materials, and thus less interesting from an economic point of view. We disregard these industries in our analysis because they are not classified as manufacturing industries.

Table 1.3 – The 10 most diverging service and manufacturing industries according to δ_U

NAF700	Industry	δ_U	# of plants
	Service Industries		
660F	Reinsurance	0.876	38
621Z	Scheduled air transport	0.434	320
671A	Administration of financial market	0.374	38
741E	Market research and public opinion pooling	0.331	1323
611B	Coastal water transport	0.326	65
602C	Cable cars and sport ski lifts	0.305	221
631A	Cargo handling	0.281	130
611A	Sea transport	0.256	173
632C	Other supporting water transport activities	0.240	392
724Z	Database Activities	0.239	654
	Manufacturing Industries		
351A	Building of warships	0.578	24
172G	Silk-type weaving	0.568	116
182E	Manufacture of women's outerwear	0.559	2248
172C	Woolen-type weaving	0.539	17
171E	Preparation of worsted-type fibers	0.537	56
274A	Precious metals production	0.518	32
181Z	Manufacture of leather clothes	0.515	97
362A	Striking of coins	0.505	24
153A	Processing and preserving of potatoes	0.487	106
221G	Publishing of sound recording	0.481	868

1.4.2 Result #2: Services are localized at shorter distances than manufacturing industries

Our second result concerns the spatial scope of localization in manufacturing and service industries. We disentangle localization from dispersion for each industry, using the distance-dependent test introduced in section 1.2.2. As localization depends on a specific threshold distance, we first propose a sorting of industries in four exclusive categories depending on the threshold we consider. We then compare how service and manufacturing industries are sorted into these categories. We put forward that a larger share of service industries are localized at short distances (below 4 km) than for manufacturing industries.

1.4.2.1 Industry Sorting

We compute our distance-dependent test for the three following thresholds: $d = 4$ km, $d = 40$ km and $d = 180$ km.[25] The significantly diverging industries[26] can then be

[25]Those values correspond respectively to the median radius of French municipalities (*communes*), employment areas and *régions*. The French continental territory is partitioned into 36247 municipalities, 341 employment areas and 21 *régions*. See chapter 5 for more details.

[26]We consider in the following a 5% level test for significant divergence.

classified in one of these four exclusive categories:

- **Case 1: f^{obs} already diverges from the reference distribution before 4 kilometers.** Our interpretation is that industries localized at such a small scale are organized around one or few highly localized clusters. We guess that these industries are the one for which face-to-face interactions and informal contacts are the most valuable.

- **Case 2: f^{obs} diverges from the reference distribution before 40 km, but not before 4 km**. Our interpretation is that these industries are spatially localized in a small number of labor markets.[27] Our prior is that these industries have some specific labor requirements that force them to locate in specific local labor markets. However, those industries do not specifically require to be clustered at shorter distances (contrary to case-1 industries).

- **Case 3: f^{obs} diverges from the reference distribution before 180 kilometers, but neither before 40km, nor before 4 km.** Our prior is that input-output linkages are the main drivers of the location choices for these industries. We consider such large-scale localization patterns because Duranton, Martin, Mayer, and Mayneris (2008) note that vertically-linked industries co-localized at such a distance. Moreover, Maurel and Sédillot (1999) argue that, in France, some industries display spatial concentration at the regional level.

- **Case 4: f^{obs} diverges from the reference distribution beyond 180 kilometers.** This case corresponds to our definition of dispersion.

We emphasize here Marshallian externalities as the main drivers of localization. The presence of local specific endowments or large indivisible facilities can also be a strong determinant of location. However, we remove from our sample any extractive industries whose location choices are the most constrained by the availability of raw materials. Moreover, we believe that underlining the differences in the spatial scope of localization can give some information on the specific mechanisms driving its location choices. For instance, enjoying technological spillovers is a sufficient, but not necessary, condition for firms to cluster in a small scale area.

It is worth noting that although this sorting does not depend on the (unweighted) index of divergence δ_U, industries with the largest δ_U are mostly case-1 industries. As an illustration, 70% of the case-1 industries belong to the upper quartile of the distribution of δ_U. Conversely, 69% of industries in this upper quartile are case-1 industries.[28] This suggests that the most diverging industries are first of all industries localized at very short distances.

[27] In France, employment areas are defined so as to minimize daily cross-boundary commuting. They more or less enclose a self-contained local labor market. The mean radius of an employment area gives thus a good benchmark for the size of a labor market.

[28] These figures stand at 69% and 73% with one considers the weighted test for localization and the weighted index δ_W.

1.4.2.2 Comparing localization and dispersion in manufacturing and service industries

Table 1.4 outlines the sorting of manufacturing and service industries in each of the four previous mentioned cases. It clearly stands out that service industries mostly belong to the first case. This means that a majority of service industries are localized at very short distances (before 4 km). On the contrary, the patterns of localization in manufacturing industries are less distinctive. A large share of manufacturing industries register localization at longer distances (between 40 and 180 km, or even beyond).

Not only do services diverge more often from randomness, but they also mostly diverge at very short distances. The spatial scope of localization for services is significantly smaller than for manufacturing industries.

Table 1.4 – Sorting of service and manufacturing industries

	Unweighted Test		Weighted Test	
	Manuf.	Service	Manuf.	Service
Non-Diverging	32%	4%	42%	18%
Localized $<$ 4km	16%	53%	18%	55%
Localized $<$ 40km	9%	13%	6%	16%
Localized $<$ 180km	18%	10%	13%	8%
Dispersed $>$ 180km	26%	20%	22%	3%

An obvious explanation for this striking difference is that service industries are mainly located in the heart of a few French big cities. It echoes the result by Arzaghi and Henderson (2008) in the case of advertising activities in Manhattan, but extents it to a bunch of business-oriented service industries. Examples of such industries are financial market administration, reinsurance or market research and public opinion pooling. These industries are among the most diverging industries found in table 1.3.

Furthermore, if not localized at short distances, service industries appear mainly dispersed. Once more the *urbanness* of services, pointed out by Holmes and Stevens (2004), is at play. However, contrary to the previous case, plants in these industries are not only located in big cities, but also in small and medium cities. Contrary to manufacturing industries, the absence of these industries from most rural areas explain why they nevertheless remain significantly diverging. The main driver for these service industries is proximity to consumers. Sewage and refuse disposal, banks, renting of automobiles, storage and warehousing are examples of such services.

1.4.3 Comparisons with alternative measures of localization

Before turning to a more detailed analysis of the patterns of location in services, we compare our results with what is obtained with a DO approach on one hand and with an EG approach on the other hand.

A comparison with the DO index

In section 1.2.4, we argue that the two main differences of our setup with the DO approach concern: 1/ the smoothing procedure in the estimation of the density distribution and 2/ the definition of the test for significance. We first test the robustness of our results to the smoothing procedure, and then compare these results with what is found with the DO approach.

Table 1.5 provides the same information as tables 1.2 and 1.4 when all the density distributions are estimated with a kernel smoothing procedure *à la* Silverman (1986) instead of a simple histogram estimator. The results are quantitatively the same and our two previous conclusions remain valid: 1/ service industries diverge more often from randomness than manufacturing industries, 2/ in most cases, this divergence of services corresponds to localization at very short distances (before 4 km).

Table 1.5 – Summary statistics
Indices computed with a smoothing procedure à la Silverman (1986)

	Unweighted Index (δ_U)		Weighted Index (δ_W)	
	Manuf.	Services	Manuf.	Services
Sorting of industries...				
Diverging at the 5% level	71%	93%	53%	83%
Non-Diverging	29 %	7 %	48 %	19 %
Localized < 4km	*20%*	*43%*	*12%*	*59%*
Localized < 40km	*3%*	*9%*	*6%*	*9%*
Localized < 180km	*20%*	*9%*	*15%*	*5%*
Dispersed > 180km	*29%*	*31%*	*20%*	*7%*
Moments of the index distribution				
Mean of δ	0.070	0.073	0.092	0.144
Median of δ	0.044	0.044	0.056	0.074
Variance of δ	0.009	0.012	0.016	0.024
Min of δ	-0.046	-0.011	-0.167	-0.094
Max of δ	0.578	0.899	0.690	0.776

We then compare the results obtained with a smoothing procedure with what is found with a DO approach. From there, the remaining difference between both approaches stands only in the way the test for significance is computed. Table 1.6 provides the share of globally localized industries (before 180 km) according to Duranton and Overman (2005)'s definition and the share of localized industries before 180 km in our setting (sum of cases 1, 2 and 3 in table 1.5). It appears that the DO approach overestimates the extent of localization in both cases by around 10%. We develop in appendix 1.7 a longer comparison of both approaches and conclude that the DO approach is upward biased, the reason why we developped this alternative test. Note however that our first conclusion remains valid when using the DO approach.

Table 1.6 – A comparison with the DO approach
Localization before 180 *km*

	DO approach	Our approach
Manuf.	53%	42%
Service	71%	61%

A comparison with the EG index

Let us compare our results (in the weighted approach) with what is found with an EG index. Table 1.7 in introduction provides moments for the distribution of the EG index computed for both manufacturing and service industries at the employment-area level. We already note that the extent of spatial concentration is much smaller for services than manufacturing industries when considering an EG index. Table 1.2 shows that the reverse result holds true with our methodology in the weighted approach. We thus argue that the EG index underestimate the extent of spatial concentration in business services.

Surprisingly enough, the test for the significance of spatial concentration originally proposed by Ellison and Glaeser (1997) is rarely computed in the literature. Table 1.7 provides the results from such a comparison. 76% of industries are found spatially concentrated (at the 5% level)[29] with an EG index computed at the employment-area level: 74% among manufacturing industries and 85% among service industries. These figures are slightly larger than our results. As noted by Duranton and Overman (2005), the EG index overestimates the number of industries which depart from a random distribution. The gap is more pronounced for manufacturing than for service industries. Hence, we believe that our distance-based approach which is not affected by the MAUP is better suited to extract meaningful information about the extent and scope of localization of service industries than the EG index.

Table 1.7 – A comparison with the EG index

		Overall	Manuf.	Services
Tests at 5% level	% diverging industries	64	58	82
	% concentrated industries (EG)	76	74	85
Tests at 10% level	% diverging industries	67	62	82
	% concentrated industries (EG)	79	77	86

Finally, the rank correlation between our weighted index δ_W and the EG index is 0.54, whereas the rank correlation between the weighted DO index of localization and the EG index is only 0.26.[30] Undoubtedly, our index is more correlated to the EG index than the DO measure. A large part of this greater correlation is due to the analogy we put forward earlier.

[29] Results for significance are obtained from a test done at a 5% level, using the theoretical variance provided by Ellison and Glaeser (1997) (p 907).

[30] These figures stand respectively at 0.70 and 0.51 when restricted to industries with strictly positive DO index of localization.

1.4.4 Robustness checks

To complete our cross-industry analysis, we test the robustness of the two previous conclusions to three arbitrary choices in our methodology: the choice of an industrial classification, of a set of active sites, and of threshold distances in the sorting of industries.

How does the industrial classification matter?

The 4-digit French industrial classification (NAF700) is less detailed for services than for manufacturing industries, as proved by the mean number of plants in each group. Even if our test for and index of divergence is comparable across industries with different industrial concentration, aggregation of different activities within the same industry could still affect the results beyond a systematic industry-size effect. Indeed, it could be the case, for instance, that some *co-localized* industries are gathered in the same item for services but in separate items for manufacturing industries. Furthermore, when considered separately these industries could display no distinctive patterns of localization, but appear as localized when considered as only one industry.

To illustrate that point, we carry out the computation of our index on a less detailed, 3-digit industrial classification (NAF220). That classification aggregates the 311 manufacturing industries into 105 *sectors*. Concerning services, we keep the 4-digit classification (NAF700), with its 96 service industries. Our prior is that such an aggregation of manufacturing industries makes manufacturing *sectors* and services *industries* more comparable. Table 1.8 provides the results.

Table 1.8 – Summary statistics: sectors *versus* industries

	Unweighted Index (δ_U)	
	Manuf. (NAF220)	Services
# sectors/industries	105	96
Sorting of sectors/industries...		
Diverging at the 5% level	88%	96%
Non-Diverging	12%	4%
Localized < 4km	*21%*	*53%*
Localized < 40km	*10%*	*13%*
Localized < 180km	*30%*	*10%*
Dispersed > 180km	*27%*	*20%*
Moments of the index distribution		
Mean of δ	0.078	0.103
Median of δ	0.054	0.055
Variance of δ	0.007	0.015
Min of δ	-0.057	-0.048
Max of δ	0.515	0.876

At the 5% level of significance, the share of diverging service industries (96%) remains larger than the corresponding figure for manufacturing industries (88%). However, the share of diverging manufacturing industries increases with sectoral aggregation, from 68%

at the 4-digit level to 88% at the 3-digit level. Hence, part of the high share of diverging services may be due to aggregation of different colocalized activities within the same 4-digit item. However, this channel cannot explain the whole difference between services and manufacturing industries.

The second result is remarkably robust to the aggregation of manufacturing industries. Service industries remain more often localized at short distances than manufacturing sectors even if aggregation increases the share of manufacturing sectors localized before 4 km from 15% to 21%.

To conclude, both results 1 and 2 hold when we use classification items providing more comparable details for service and manufacturing activities. Co-localization of different service activities within the same 4-digit NAF700 items, though present, cannot explain the whole difference between services and manufacturing industries we put forward earlier.

How does the set of active sites matter?

The set of active sites is the whole set of plant locations, whatever the industry this plant belongs to. However, manufacturing and service plants could make very different location choices, due to differences in the land intensity of their activities for instance. In this section, we consider that plants in service industries can only be allocated to active sites where a service plant is actually settled, and respectively, that manufacturing plants can only be allocated to active sites where a manufacturing plant is actually located. We thus consider two different reference distributions. The reference distribution for service industries consists in the density distribution of bilateral distances between all pairs of *service plants* only; and respectively for manufacturing. In other words, we compare the location patterns for a given service (resp. manufacturing) industries to the overall location patterns of service (resp. manufacturing) plants: a *within-group* comparison.

Table 1.9 – Summary statistics: different sets of active sites

	Unweighted Index (δ_U)	
	Manuf.	Services
Sorting of industries...		
Diverging at the 5% level	77%	95%
Non-Diverging	23%	5%
Localized < 4km	*30%*	*44%*
Localized < 40km	*11%*	*8%*
Localized < 180km	*15%*	*14%*
Dispersed > 180km	*21%*	*29%*
Employment-weighted moments of the index distribution		
Mean of δ	0.093	0.090
Median of δ	0.048	0.047
Variance of δ	0.016	0.014
Min of δ	-0.121	-0.080
Max of δ	0.610	0.865

Table 1.9 presents the same results as tables 1.2 and 1.4 in this new set-up. Both results 1 and 2 hold, though in a lesser extent. Service industries remain more often diverging (95%) than manufacturing industries (77%). Divergence mostly occur at short distances (before 4 km). The previous results do not completely rely on the very different location choices of service and manufacturing industries. Even in an *within-group* perspective, services present more uneven patterns of location than manufacturing industries. These patterns of location are characterized by a stronger tendency toward localization.

How do the threshold distances matter for the sorting of industries?

The sorting of industries we use in section 1.4.2.1 is pretty conservative. We define as dispersed an industry whose distribution diverge significantly from the reference distribution beyond 180 km. We test for the robustness of our results by decreasing all threshold distances by 25% (upper panel of table 1.10), or increasing these distances by 25% (lower panel of table 1.10). None of our previous results are significantly affected by this change. More than half of service industries are localized at very short distances (between 3 and 5 km). The same figure stands only around 15% for manufacturing industries. Manufacturing industries appears mostly localized at very long distance or even dispersed.

Table 1.10 – Sorting: sensitivity to the threshold distances

	Unweighted Test		Weighted Test	
	Manuf.	Service	Manuf.	Service
Non-Diverging	32%	4%	42%	18%
	25% decrease in the threshold distances			
Localized $<$ 3km	17 %	53 %	18 %	58 %
Localized $<$ 30km	6 %	11 %	5 %	12 %
Localized $<$ 135km	16 %	7 %	11 %	6 %
Dispersed $>$ 135km	29 %	24 %	24 %	5 %
	25% increase in the threshold distances			
Localized $<$ 5km	16 %	53 %	15 %	58 %
Localized $<$ 50km	10 %	14 %	8 %	15 %
Localized $<$ 225km	24 %	18 %	23 %	2 %
Dispersed $>$ 225km	19 %	11 %	13 %	7 %

1.5 Within-industry results

In this section we consider localization patterns *within* industries. We focus on two important questions. The first one, initiated by Holmes and Stevens (2002), concerns the relationship between plant size and localization. The second one, originated in Dumais et al. (2002)'s work, is about the dynamics of localization.

1.5.1 Result #3: Large plants in service industries are the main drivers of localization

A pervasive question in the literature concerns the localization patterns of the largest plants in comparison with the localization patterns of the smallest ones (see for instance Holmes and Stevens, 2002; Duranton and Overman, 2008). There are several reasons why the largest plants may be more localized. Okubo, Picard, and Thisse (2008) embed a Melitz-type model of monopolistic competition with heterogeneous firms in a new economic geography framework. Given an exogenous distribution of plant productivity, they show that, for some intermediate values of transport costs, the most productive plants, and thus the largest ones, are concentrated in one region: a sorting effect. Holmes and Stevens (2002) argue that, in a dynamic setting, if plants benefit from any kind of localization economies, they could be more productive, and hence, grow faster in areas where an industry is concentrated than outside such areas.

We reconsider this question for both manufacturing and service industries. A first approach is to compare our weighted and unweighted indices in table 1.2. This comparison suggests that the largest plants are more unevenly located than the other plants within most service industries. The same is not observed for manufacturing industries. Indeed, the weighted index strongly depends on the location choices of the largest plants. It stands out in table 1.2 that the distribution of the weighted index (δ_W) for services is shifted to the right in comparison with the distribution of the unweighted index (δ_U).[31] Such a shift in the distribution of δ is not observed for manufacturing industries.[32] The mean and median values of δ_W are 1.5 times larger than the mean and median values of δ_U concerning services.

The previous comparison is a first proof that the largest plants in services display uneven patterns of location. However, the ultimate question is to understand whether or not these plants are more localized than the overall plants within those industries? To further investigate this question, we consider a slightly different approach, inspired from Duranton and Overman (2008). For each industry with more than 100 plants, we select the 10% largest plants. We first compute the distribution of bilateral distances between these largest plants in each industry and assess its divergence against an industry-specific reference distribution. In this case, the reference distribution is the (previously-defined) observed distribution of bilateral distances between all pairs of plants in the industry under scrutiny. In order to assess whether this divergence is significant or not, we randomly allocate these largest plants across the whole set of active sites for this industry. In other words, we consider that, for each industry, the potential location choices of the largest plants is the set of sites where a plant from the industry is located whatever its size. This is the simplest way to handle an *within-industry* comparison in the location choices of the largest plants.

[31] The shuffling of service industries within the distribution is quite limited. The rank correlation between δ_U and δ_W stands at 0.67 and the Pearson correlation at 0.84 for services.

[32] The rank correlation between δ_U and δ_W stands at 0.73 and the Pearson correlation at 0.88 for manufacturing industries.

As previously, we repeat this simulation step 1000 times and compute the 5% level test of significance using the empirical distribution of divergences obtained through these simulations. We are able to compute a test for divergence (on the whole set of distances) and distance-dependant tests for localization. Broadly speaking, we study whether the number of large plants in the vicinity of a large plant is larger than the number of small plants: an evidence of the localization of large plants.

Table 1.11 provides the share of service and manufacturing industries for which the largest plants appear significantly diverging (upper part of the table), and detailed the specific threshold distance[33] at which localization occurs (lower part of the table). A striking feature is the high share (63%) of service industries whose largest plants are localized at very short distances (before 4 km). On the contrary, within manufacturing industries, largest plants are most of the time non-diverging.

Table 1.11 – Location patterns of the largest plants

	Manuf.	Services
# industries	186	89
Sign. Diverging	43%	75%
Non-Diverging	57%	25%
Localized $<$ 4km	25%	63%
Localized $<$ 40km	4%	8%
Localized $<$ 180km	3%	1%
Dispersed $>$ 180km	11%	3%

Using a completely different approach, Holmes and Stevens (2002) find on US data for service and manufacturing industries, that plants located in areas where an industry is concentrated are on average larger than plants outside such areas. Holmes and Stevens (2002) emphasize their result for manufacturing industries. However, the same holds true in their study for services. For instance, in the 5% most specialized census regions, manufacturing plants are 35% larger than their average size in the US. At the same time, plants in FIRE (finance, insurance and real estate) and business service are respectively 29% and 20% larger.[34] Our results give support to Holmes and Stevens (2002)'s findings for services but only partially confirm their findings for manufacturing industries. Moreover, we go one step further by showing that the largest plants are not only located in specialized areas, but also surrounded by other large plants. Duranton and Overman (2005) also note for UK that the localization of the largest plants is not as widespread as suggested by Holmes and Stevens (2002) for manufacturing industries.

At least two non-exclusive explanations can be given to this marked difference. Either sorting/selection of the most productive plants is stronger in services than in manufacturing industries. Or, in a dynamic setup, if service plants enjoy stronger localization economies, they can grow faster and larger. In their meta-analysis of agglomeration economies, Melo,

[33]We keep here the thresholds from section 1.4.2.

[34]See their table 7 page 689.

Graham, and Noland (2009) indeed suggest that agglomeration economies may be stronger for service industries.

1.5.2 Result #4: For most service industries, new plants reduce localization whereas exiters reinforce it

All the results presented so far are static, given for the year 2005. However, localization patterns observed in a given year are the outcome of a complex spatial industrial dynamics, as initially highlighted by Dumais et al. (2002). Understanding how industrial dynamics interacts with location choices is particularly relevant from a policy perspective. For instance, the success of cluster policies heavily depends on the possibility for policy makers to curb this dynamics. In this section, we detail the dynamics of localization patterns in manufacturing and business service industries in France over the period 1996-2005.

Plant turnover and the dynamics of localization

We study the impact of plant creations and destructions on industry patterns of localization. In a given year, we can split up the stock of plants into three categories: new plants, continuing plants, and future exiters. The stock of bilateral distances within an industry can accordingly be divided into six categories: 1/ between new plants only, 2/ between continuing plants only, 3/ between exiters only, 4/ between new and continuing plants, 5/ between new plants and exiters, and finally, 6/ between continuing plants and exiters.

Within an industry, the flow of new plants modifies the density distribution of bilateral distances through its impact on categories 1, 4 and 5. Respectively, the flow of exiters impacts on categories 3, 5 and 6. The aim of this subsection is to detect whether, within each industry, the density distribution of bilateral distances in each of this two sub-groups significantly differ from the overall distribution of bilateral distances in the current year. In other words, we test how new plants or exiters impact on the overall distribution of bilateral distances within an industry.[35]

More precisely, we consider the impact of new plants (plants created between 1996 and 2000), and exiters (plants disappearing between 2000 and 2005) on the density distribution of bilateral distances in each industry for the year 2000. In this section, due to data availability, all plants are located at the cityhall's spatial coordinates of the municipality they are located in. Recall that French municipalities are very small spatial units, and such an error in location remains of second order concerns. To use our previous approach, we first compute the distribution of bilateral distances involving at least one new plant (exiter) as follows:

$$f^{new}(d) = \frac{1}{(n_e(n_e-1)/2 + n_e(n-n_e))} \sum_{i=1}^{n_e} \sum_{j=i+1}^{n} \mathbb{I}_{d \leq d(i,j) < (d+1)}, \qquad (1.9)$$

with n the total number of plants in the industry (new plants, continuing plants and exiters) ranked so that the first n_e plants are new. As previously, $d(i,j)$ stands for the distance

[35] Duranton and Overman (2008) also study the one-year dynamics of plants and its impact on localization. They use a slightly different approach by considering separately category 1 and then categories 4 and 5.

between the plants i and j. d lies between 0 and 1109. With this notation, we register $n_e(n_e - 1)/2$ bilateral distances between new plants only (category 1), and $n_e(n - n_e)$ bilateral distances between new and other plants (categories 4 and 5). We symmetrically define a density distribution for exiters.

We then compute the divergence between this distribution (f^{new}) and the overall distribution of bilateral distances within the industry under scrutiny for the year 2000 (f^{obs}_{2000}). Note that the reference distribution in this setup is again industry-specific. Finally, we test for the significance of this divergence by randomly reallocating plants (whatever its type: new, continuing or exiting) across all sites occupied by the industry in the year 2000. We draw 1000 simulations for each industry.

To avoid noisy variations, we only consider industries with more than 100 plants in 2000. We finally work with 182 manufacturing industries and 71 service industries.

How does plant turnover impact on localization?

We sort new plants and exiters in the four cases described in section 1.4.2. Table 1.12 outlines the results for manufacturing and service industries respectively.

Table 1.12 – Turnover in manufacturing and service industries

	New plants		Exiters	
	Manuf.	Service	Manuf.	Service
Non Diverging	46 %	14 %	54 %	8 %
Localized $<$ 4km	2 %	8 %	3 %	1 %
Localized $<$ 40km	7 %	18 %	12 %	31 %
Localized $<$ 180km	16 %	27 %	8 %	41 %
Dispersed $>$ 180km	29 %	32 %	23 %	18 %

First, new plants and exiters clearly more often impact on the location patterns of service industries than on the location patterns of manufacturing industries. The distribution of bilateral distances involving any new or exiting plants is not significantly diverging from the distribution of overall bilateral distances for about 50% of manufacturing industries, whereas this figure is only around 10% for service industries.

Second, new plants and exiters are rarely localized at very short distances (before 4 km). It suggests that entries and exits take place outside existing clusters. Otherwise, entrants and exiters would be colocalized with continuing plants of those clusters which would increase the divergence of their distribution at short distances. Nevertheless, this result is not sufficient to conclude that entrants decrease localization whereas exiters increase it. We need to take into account the distance at which the whole industry appears localized.

We thus want to disentangle industries with a tendency to localization from industries with a tendency to dispersion. For each industry, we are able to determine whether new plants and exiters reinforce the industry localization or dispersion by comparing the sorting of new plants and exiters (table 1.12) to the sorting of overall plants determined in the previous sections (table 1.4). For instance, for a given industry, if new plants are localized at a distance shorter or equal to the one determined for overall plants, we can assert that new plants increase localization.[36] Table 1.13 outlines the results for manufacturing and service industries.

Table 1.13 – Increasing localization or increasing dispersion?

		Service			Manuf.		
		Entrants			Entrants		
		↗ dispersion	no effect	↗ localization	↗ dispersion	no effect	↗ localization
Exiters	↗ dispersion	3%	4%	15%	4%	6%	10%
	no effect	1%	6%	1%	9%	37%	9%
	↗ localization	59%	4%	6%	19%	5%	1%

Clear dynamic location patterns can be read on the diagonals of table 1.13. Regarding manufacturing industries, in a large number of cases (37%), both new plants and exiters have no significant impact on the dynamic patterns of location. This figure stands only at 6% for services. But, within 59% of service industries (against 19% only for manufacturing industries) new plants tend to reduce localization whereas exiters reinforce it. However as new plants are more numerous than exiters for most industries,[37] we conclude that for those industries plants are more and more scattered.

Dumais et al. (2002) put forward using plant-level data for the US that spatial concentration is rather stable across time though a high degree of spatial mobility of each industry. They further show that new plants tend to decrease localization as they mostly locate outside existing clusters. On the contrary, the probability of failure is higher outside clusters, and thus exiters tend to reinforce localization.[38] Our results comfort their findings but, once again, this tendency is far much stronger for service industries than for manufacturing industries.

1.6 Conclusion

This paper studies the location patterns of business-oriented services and manufacturing industries in France considering space as continuous.

The first contribution of the paper is to develop a new test for localization fully comparable across industries with different numbers of plants and employment distribution across

[36] As previously, we look to the three following distances 4, 40 and 180 km.

[37] For $3/4$ of industries under scrutiny the number of entrants exceeds the number of exiters and for half of them the number of entrants is more than 15% greater than the number of exiters.

[38] Using the same methodology, Barrios, Bertinelli, Strobl, and Teixeira (2005) find the same results for Ireland and Portugal.

plants. It is useful when comparing service and manufacturing industries. We first prove that the approach suggested by Duranton and Overman (2005) suffer from a systematic upward-bias and that this bias increases with the number of plants in the industry. Our methodology is free from such a bias. The intuition of the test is to assess whether the divergence between the observed density distribution of bilateral distances between pairs of plants within an industry and a well-defined reference distribution is significantly larger than what would prevail under a purely random allocation over space of plants in that industry. Building on this measure of divergence, and following insights from Ellison and Glaeser (1997), we also suggest an index of divergence with two desirable properties: being insensitive to the MAUP and independent of the industrial concentration.

The second contribution of this paper is to highlight some distinctive locational features of services in comparison with manufacturing industries. We show that a distance-based approach is better suited than the traditional EG employment-based index of spatial concentration to extract meaningful conclusions concerning the location patterns of services. We highlight four main results: 1/ service industries diverge more often from randomness than manufacturing industries, 2/ a majority of diverging service industries are localized at very short distances (before 4 km) whereas manufacturing industries appear in majority localized at larger distances or even dispersed , 3/ the largest plants in service industries are even localized (at short distances) in comparison with the overall location patterns of plants in their own industry, 4/ within most service industries, new plants reduce localization whereas exiters reinforce it, which is observed for only one fifth of manufacturing industries.

We finally check the robustness of our two first results against three modifications. First, we re-aggregate manufacturing industries in broader manufacturing sectors more comparable in terms of plant numbers with service industries. Second, we consider alternative random benchmarks in order to evaluate localization. Finally, we modify the thresholds distances used to establish our results. Our results remain valid to these modifications.

1.7 Appendix to chapter 1: On the consistency of the test for divergence

In this section, we give a brief overview of the test for localization developed by Duranton and Overman (2005) and we empirically prove that this test is systematically biased in small sample. Furthermore, we show that the size of the bias increases with the number of plants in the industry. On the contrary, we show that our test for divergence developed is not biased.

A quick overview of the DO approach

The underlying idea of the test for localization suggested by Duranton and Overman (2005) is to compare the density distribution of bilateral distances between plants within an industry to the density distribution of bilateral distances within a hypothetical industry with the same number of plants randomly allocated across all active sites.

Their methodology can be reviewed in three major steps. The interested reader will find more exhaustive details in the original paper by Duranton and Overman (2005).

1. They build the observed distribution of bilateral distances between all plants in a given industry. This density distribution is estimated by a Gaussian kernel-smoothing estimator with a rule-of-thumb bandwidth *à la* Silverman (1986).

2. They build counterfactuals to which the observed density distribution is compared. These counterfactuals are drawn from simulations. These simulations consist in randomly reallocating plants of the considered sector across all active sites. The set of active sites are locations where a manufacturing plant is currently located, whatever its sector. For each simulation, they are able to compute the density distribution of bilateral distances. They draw 1000 simulations.

3. Then, from this set of density distributions, they build a local and a global confidence interval. Note that Duranton and Overman (2005) do not consider the entire range of bilateral distances ($[0 - 1109]$ km in our case) but only the range of distances until the median ($[0 - 392]$ km). Local confidence intervals are then build by dropping from the sample of simulated densities the five percent of the lowest and greatest observations *at each distance*. Taken together, a large fraction (well above 5% percent) of the simulated densities are then dropped from the sample over the range $[0 - 392]$ km. These local confidence intervals are then too restrictive, and do not allow making any statements about the global location patterns of a sector. It is the reason why they define a global confidence interval. To build this interval, they search for the local confidence threshold, common to each distance, so that only 5% of all randomly generated densities are dropped from the sample.[39]

[39]The precise construction of global confidence intervals requires a complicated step-by-step procedure presented in details in Duranton and Overman (2005) and Klier and McMillen (2008).

Finally, an industry is said to be globally localized (at a 5% confidence level) if the observed density distribution of bilateral distances hits the upper band of the global confidence interval for at least one distance over the range $[0 - 392]$ km. On the contrary, a sector is said to exhibit global dispersion (at a 5% confidence level) if the observed density of bilateral distances never hits the upper band and hits the lower band of the confidence interval for at least one distance over the range $[0 - 392]$ km.

The local test for significance is not questioned here. We only consider the global confidence interval. Obviously, their method would define the correct envelope if they were able to compute the whole set of possible simulations. Nevertheless, we argue that it suffers from an upward-bias in small samples. This small-sample upward bias is then of primary interest because computing the whole set of simulations is not tractable, the reason why Duranton and Overman (2005) compute only 1000 simulations.

Recall that if N is the number of active sites and N_i the number of plants in the industry i under scrutiny, then $C_N^{N_i}$ is the total number of feasible simulations in the population. Thus, 1000 simulations represent only a very small sample from the population. Moreover, the larger the industry, the smaller the sample. Indeed, $C_N^{N_i}$ is firstly upward-sloping and hits its maximum for $N_i = \frac{N}{2}$, well above the number of plants in any industry.

We note two other drawbacks in the construction of $\bar{\bar{K}}(d)$, the envelope of the global confidence interval. First, $\bar{\bar{K}}(d)$ does not belong to the space of density distributions. Thus, it leads to compare two different kinds of mathematical objects. Second, whereas 5% of the random simulations lie strictly above $\bar{\bar{K}}(d)$, there may not be 95% of them strictly below. Indeed, $\bar{\bar{K}}(d)$ is an envelope of simulations, and then some simulations are part of this envelope. In other words, whereas usual confidence bounds have a zero measure, this is not the case here.

In what follows, we prove that the method proposed by Duranton and Overman (2005) really leads to an upward-biased test for global localization in small samples and, more importantly, that this bias increases with the number of plants in the industry. The main goal of this paper is to propose an alternative unbiased test for localization (see section 1.2.2).

Empirical evidence of an upward-bias in the DO test for global localization

If the test proposed by Duranton and Overman (2005) were unbiased, any randomly distributed industry should lie above its $\bar{\bar{K}}(d)$ with a 5% chance. It is worth noting that $\bar{\bar{K}}(d)$ depends only on the number of plants in the industry (say, N_i) in the unweighted DO approach.

We put forward the existence of a bias in the unweighted DO test using the following procedure. We compute confidence intervals corresponding to 407 randomly-distributed industries with numbers of plants, from 11 to 28,909 plants.[40] For each number of plants, we compute $\bar{\bar{K}}(d)$ using the DO method, except that we draw only 500 simulations (instead of 1000).[41]

[40]These numbers of plants correspond to those of the French industries considered in the main text.
[41]This limitation is imposed by computer capacities.

Then, for each of the 407 confidence intervals,

1. we create a fictive industry with the corresponding number of plants by randomly allocating them across active sites.

2. we then test whether this randomly-distributed industry is localized or not using the corresponding confidence interval.

We repeat these two steps 500 times and count the number of randomly-distributed industries that appear as localized. If the DO test were unbiased, this number should, on average, be equal to 25 (5% of 500) for each number of plants under scrutiny. Moreover, it should be independant of the number of plants.

Figure 1.2 – Bias of the DO test for global localization

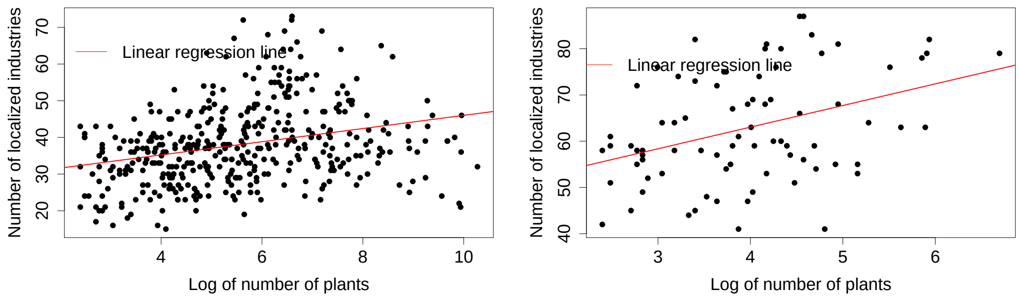
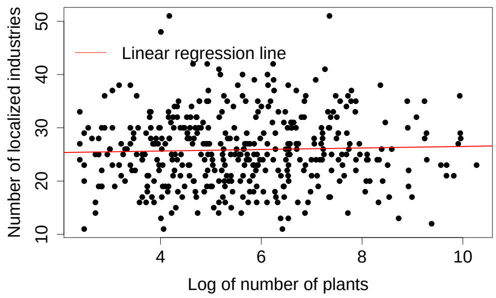


Figure 1.2 (left-hand graph) shows that none of this two assertions is supported by our test. The average number of localized industries is 38 (instead of 25). The linear regression of the number of localized industries on the log of the number of plants in the industry gives a coefficient equal to 1.8. This coefficient is significant at the 1% level and the R-squared of the regression equals to 8.4%. Consequently, we prove that the DO global localization test is upward-biased in small samples and that the bias is positively correlated with the number of plants in the industry.

One could argue that our results are due to the small number of simulations (500 instead of 1000) used to build the global confidence interval. For the smallest industries, we are able to compute 1000 simulations, and 1000 randomly-allocated industries to test for the bias. Figure 1.2 (right-hand graph) shows the number of randomly-distributed industries found to be localized as a function of the number of plants in the industry for this subset of small industries (80 industries).

If the DO test were unbiased, the average number of localized industries should be 50 (5% of 1000), whatever the number of plants in the industry. We find that the average value is 63, and that this value is again increasing with the number of plants in the industry (slope of 4.7, significant at 1% level). We find on this subset of randomly-distributed industries that moving from 500 to 1000 simulations slightly reduces the bias in the DO

Figure 1.3 – Results for our test of divergence



test, but does not eliminate it. This confirms that the DO test is asymptotically consistent, but systematically upward-biased in small samples.

On the consistency of our test for divergence

We show in this section that our test for divergence is unbiased. We repeat the previous test.

Figure 1.3 shows the number of randomly-distributed industries found to be diverging with our test as a function of the number of plants for the 407 previous defined industries. We compute the test using only 500 simulations. The mean value stands at 26 (instead of 25 (5% of 500),[42] and the slope of the regression is 0.14 (non significant at any usual level). This confirms that our test is neither biased, nor dependent of the number of plants in the industry.

Does the bias matter?

The bias we put forward previously may appear relatively small. The size of the bias with 1000 simulations is, on average, 1 point but it is almost 3 points for industries with the largest number of plants. However, we cannot know in advance how many industries are going to be affected. Even if the bias is low, it may matters for a lot of industries. Actually, the bias affects the results of all industries whose observed density distribution lies slightly above the upper bound of the confidence interval.

To give a concrete example, we compute on our data both the DO test and our test for localization. Both tests are computed at the median distance across all active sites in our sample, 392km. Moreover we use in both cases the DO smoothing procedure in the estimation of density distributions, so that only the test will affect the results.

Unsurprisingly, the DO test overestimates the number of industries considered as localized. On the whole sample, the DO test concludes that 63% of industries are localized whereas our test concludes to only 55%. The gap between the results of the two tests dramatically increases with the number of plants in the industry. For industries with more than

[42]The discrepancy between the theoretical value (25) stands within the confidence interval of the estimated value.

Table 1.14 – Comparing both tests for localization (at 392 km)
Industries sorted by number of plants

	All industries	10-100 plants	100-500 plants	$\geq$ 500 plants
Number of industries	407	132	131	144
Localized with the DO test	63%	36%	63%	87%
Localized with our test	55%	34%	56 %	72%

Notes: Our test is the 5%-level distance-dependent test introduced in section 1.2.2 where the distance equals to 392 km. This distance corresponds to the median distance across all active sites in our sample.

500 plants, the DO test concludes that 87% of indutries are localized whereas this figure stands at only 72% with our unbiased test. This issue may be particularly accurate when broad industrial classification are used. In that case, most industries will contain more than 500 plants. Finally, the fact that the share of localized industries increases with the number of plants with our test is an economic effect, and not any more a statistical problem.

1.8 Complementary tables

Table 1.15 – The 10 most diverging service and manufacturing industries according to δ_W

NAF700	Industry	δ_W	# of plants
	Service Industries		
660F	Reinsurance	0.785	38
621Z	Scheduled air transport	0.538	320
671C	Security broking and fund management	0.452	1625
741E	Market research and public opinion pooling	0.450	1323
671A	Administration of financial markets	0.449	38
602C	Cable cars and sport ski lifts	0.381	221
660A	Life insurance	0.381	592
744B	Planning, creation and placement of advertising activities	0.379	7379
721Z	Hardware consultancy	0.374	7488
724Z	Database Activities	0.361	654
	Manufacturing Industries		
172G	Silk-type weaving	0.609	116
362A	Striking of coins	0.544	24
221A	Publishing of books	0.542	1360
221G	Publishing of sound recordings	0.531	868
171E	Preparation of worsted-type fibers	0.527	56
181Z	Manufacture of leather clothes	0.518	97
221E	Publishing of journals and periodicals	0.512	2182
351A	Building of warships	0.505	24
172C	Woolen-type weaving	0.500	17
286A	Manufacture of cutlery	0.480	153

CHAPTER 2

Agglomeration economies and firm productivity: Estimation from French individual data[1]

2.1 Introduction

The uneven distribution of economic activity across space, and especially the spatial concentration of some specific industries, is an old and well-established empirical fact, observed across various countries.[2] Some activities can be constrained in their location choices by the presence of natural or man-made endowments (see Ellison and Glaeser, 1999). However, it cannot be the sole explanation. The persistence of densely populated clusters of activity requires that firms benefit from it. It has led economists to acknowledge the existence of *agglomeration economies*, which exist as soon as an individual's productivity rises when he or she is close to other individuals.

Agglomeration economies may be *pure externalities*, as in the case where productivity rises from being able to learn from or imitate a neighbor. These agglomeration economies can also work entirely within the market. If a supplier and a customer get closer, they may become more productive only by eliminating some kind of transaction costs, but there is no obvious externality. Since Marshall (1890)'s work, agglomeration economies are understood as a way to reduce any moving costs for goods, workers or ideas (see Glaeser, 2008). However, agglomeration can also induce diseconomies, due to higher costs for immobile inputs, congestion or fiercer competition on input and output markets (see Melitz and Ottaviano, 2008, for instance). At the end, the net impact of agglomeration on productivity remains an empirical question. In this paper, we investigate how agglomeration affects average firm productivity by exploiting very detailed individual level datafiles for France.

Agglomeration economies can affect productivity in a myriad of ways that are difficult to disentangle (see Duranton and Puga, 2004, for an exhaustive overview)[3]. In front of such

[1]This paper is a joint work with Yoann Barbesol (INSEE-DEEE), published as Barbesol and Briant (2009).

[2]See chapter 1 for France, Duranton and Overman (2005) for the United Kingdom or Ellison and Glaeser (1997) for the US.

[3]Duranton and Puga (2004) provide an overview of the mechanisms through which proximity affects productivity under the headings: *sharing*, *matching* and *learning*. Sharing inputs, be it raw materials or labor, sharing new ideas, or sharing risks is the first way to explain the incentive for firms to cluster. Secondly, matching job vacancies and job seekers is easier on thick labor market. The reduction in the cost of hiring workers and the availability of a wider range of specialized skills on the local labor market are indeed other

complications, economists have paid more attention to the estimation of total net effects of agglomeration on productivity and have distinguished between *urbanization economies* and *localization econmies*. In the first case, firms benefit from the overall size of their market, regardless of the identity of their neighbors. In the second case, firms benefit from the closeness of neighbors operating in the same industry. Of course, these two categories are not mutually exclusive.

In this paper, we estimate the magnitude of urbanization and localization economies on individual firm productivity, using a large panel dataset on French firms in the manufacturing and service industries for the years 1994 to 2004. French individual data on firms are very rich and well-suited to study that question. Our dataset is build upon administrative data files on firm tax declarations, providing us with a rich array of individual firm characteristics. Our estimation strategy relies on a two-step approach. In the first one, we estimate individual firm productivity while controlling for the quality of its labor force and some sector-specific unobservables.

We then explain spatial disparities in average firm productivity by proxies for urbanization and localization economies. Once controlled for a bunch of location-specific characteristics, the major part of those disparities is explained by differences in the density of total employment, which captures urbanization economies. We find an elasticity of average firm productivity to employment density about 0.02 in our preferred specification. This result proves to be robust to the individual productivity estimation technique, to the level of sectoral aggregation and to the geographical scale. The economic effect is sizable when compared with the annual average productivity growth of French manufacturing firms in the period under scrutiny. This elasticity is also in line with values found in the international literature on the subject.

We also find evidence for localization economies, as we observe that the more concentrated in a given area an economic activity, the higher the average firm productivity engaged in that kind of business. Finally, we show that there exists a quite strong heterogeneity across industries in the elasticity of productivity with respect to employment density and specialization.

The paper is built as follows. The next section surveys the related literature. The third section introduces our estimation strategy. Then, we present our proxies for agglomeration economies and their motives. The last two sections give the baseline results and provide various robustness checks. The ultimate section concludes.

arguments in favor of the agglomeration of firms. Finally, technological spillovers, be it the share of ideas or good practices, which require proximity and face-to-face contacts, are also believed to come into play when firms cluster together geographically.

2.2 Related Literature

In this section, we review the results on the magnitude of urbanization and localization economies on individual firm productivity.[4]

Henderson (2003) is the first to introduce in a plant-level production function some proxies for agglomeration economies. His dataset consists in a non-exhaustive panel of plants in the US, observed every five years in the machinery and high-tech sectors. For each plant, the author constructs two proxies for agglomeration economies. The number of plants in the same sector and the same Metropolitan Statistical Area (MSA) is assumed to capture localization economies, whereas the overall number of plants in the same MSA, outside its own sector, is assumed to proxy for the diversity of local economic activity (or urbanization economies). As his dataset contains pieces of information on value-added, capital and employment for each plant, he is able to run an estimation of the production function at the plant level and to directly introduce these proxies into the estimation. He further introduces individual fixed effects in order to capture any unobserved firm characteristics. These fixed effects then prevents the estimation from being biased due to missing variables. He finds evidence for localization economies as the number of other own industry plants impacts positively on firm productivity in high-tech sectors but not in machinery ones. He also finds that single-plant firms benefit from and generate more external benefits than multi-plant firms.

Cingano and Schivardi (2004) use a two-step approach on Italian data. They first estimate a firm-level production function and calculate individual firm total factor productivity (hereafter TFP). Then, they compute a local sectoral productivity as the employment-weighted average of individual firm TFPs. They find that specialization enhances local sectoral productivity growth whereas diversity does not.

In a parallel line of research, Wallace and Walls (2004) consider the role of agglomeration economies in the production decisions of firms in the high-tech computer clusters in the US. Their main objective is to disentangle the effects of external agglomeration economies from scale economies internal to multi-unit firms. Following Henderson (2003), they first introduce in a firm-level production function some firm-specific proxies for internal and external network relationships, expected to capture localization externalities. Wallace and Walls (2004) further allow the coefficients of the production function (the elasticities of output to labor and capital) to vary with the economic environment of the firm. They conclude that these localization proxies significantly impact on the choice of technology by the firm.

Moretti (2004) studies how firms benefit from the existence of a large stock of human capital in their vicinity. He incorporates in a plant-level production function the share of college graduate in the city where the plant is located. He finds that the output of plants located in cities experiencing a larger increase in the share of collage graduates rises

[4]In the introduction of this dissertation, we have briefly presented two alternative approaches to quantify the magnitude of urbanization and localization economies: the local employment growth regression and the estimation of wage (or land rent) equations.

more than the output of plants in cities experiencing a smaller increase in their stock of human capital. He further shows that the output of plants is more responsive to an increase of human capital stock in economic-related activities, as defined by input-output flows, technological specialization or the frequency of patent citations.

In a recent study, Combes et al. (forthcoming) investigate the magnitude of the simultaneity bias in the estimation of urbanization economies that they label the 'endogenous quantity of labor'.[5] If a place makes firms more productive, it will induce some inward migration, leading to more agglomeration. In this case, the causality runs from productivity to agglomeration. Endogeneity due to reverse causality or simultaneity has to be controlled for. Following a two-step approach, they compute a local measure of productivity, using the same individual firm dataset as ours, that they regress on local employment density. In order to correct for the simultaneity bias, they use some historical and geological instruments for employment density. They conclude to the existence of a small bias, supporting the idea that the causality indeed runs from agglomeration to productivity. The elasticity of productivity to employment density ranges from 0.025 to 0.05 depending on their specification. The next section details the two-step approach we also use in this study.

2.3 Estimation strategy and econometric issues

In a first step, an individual firm productivity is estimated. Then, a location- and sector-specific average firm productivity, named cluster productivity, is computed. Disparities in cluster productivities are explained by proxies for agglomeration economies in a second step[6].

2.3.1 First step: estimating individual firm productivity

In the first step, productivity is computed for each firm, while controlling for the quality of its labor force and sector-specific (unobserved) determinants of productivity. Firm

[5]In their study of worker wages, they also deal with another source of bias: the 'endogenous quality of labor'. It is due to the spatial sorting of workers according to some unobservable determinants of individual productivity. In this case, some spatial disparities in productivities occur but without the existence of any externalities. To correct for this bias, they use individual fixed effects and identify their model on inward and outward migrations. Such a strategy is much more difficult to implement in the case of firms whose location changes are less frequent and more difficult to track.

[6]In the literature, an alternative one-step strategy is sometimes used, where proxies for agglomeration economies are introduced in a firm- or plant-level production function, along with individual fixed effects. Henderson (2003) uses such a one-step strategy. In his setup, identification relies on time variation in individual and cluster variables. It thus requires a long time spell of observations. Such a one-step estimation strategy is difficult to implement with our dataset for two reasons. First, our time spell is rather short: 10 years from 1994 to 2004. During this period, variations in agglomeration variables, especially urbanization variables are small. Second, the estimation of industry-specific elasticities for inputs and, at the same time, the introduction of individual fixed effects in the regression proves to be difficult to implement. So, we settle for a two-step strategy to estimate the magnitude of agglomeration economies. Martin, Mayer, and Mayneris (2008) provide results on French data using the one-step strategy.

total factor productivity is predicted as the residual of a Cobb-Douglas production function estimation:[7]

$$log(VA_{it}) = c_{st} + \theta_{multi} + \alpha_s log(L_{it}) + \beta_s log(K_{it}) + \sum_{q=2}^{3} \delta_{qt} sh_{iqt} + u_{it}, \quad (2.1)$$

where

- i indices the firm, s the sector of main activity of the firm and t the time.
- L_{it} is a measure of employment. In our data, employment is measured as the number of working hours.
- sh_{iqt} is the share of hours worked by employees of skill group q. Workers are divided up into groups according to their qualifications. This aims at controlling for the quality of the labor force. Using the French occupation classification, we set up three categories of skills: (Q3) for highly skilled workers (engineers, technicians and managers), (Q2) for skilled workers (skilled blue and white collars), finally (Q1) for unskilled workers, interns and part-time workers.
- K_{it} is a measure of the capital stock. In our dataset, this measure consists in the book value of tangible and intangible non-financial assets.
- θ_{multi} is a dummy equal to 1 if the firm controls more than one plant.
- c_{st} is a sector- and time-specific fixed effect. It captures any sector- and time-specific determinants of productivity such as: 1/ a sector-specific price index for the value-added,[8] 2/ a sector-specific age and depreciation rate of the capital stock,[9] 3/ any sector-specific macroeconomic shocks likely to affect value-added and input choices.

Labor and capital elasticities are supposed to be sector-specific. Hence, we make the assumption that the technology of production is the same for all firms in the same sector. We do not constrain the production function to have constant returns to scale. In other words, any increasing returns to scale internal to the firm are controlled for, and do not corrupt our measure of productivity.

A sector is defined as an item of the 3-digit French industrial classification (NAF220). More details about the data are provided in appendix A. Finally, note that, at this point,

[7] See for instance Aubert and Crépon (2003) or Crépon, Deniau, and Pérez-Duarte (2002) for other examples on French data.

[8] Indeed, in the production function, the dependent variable is the real value-added of the firm, which is equal to the observed nominal value-added deflated by a sector-time specific price index. This latter term is captured by our fixed effect.

[9] Our book value measure of capital is imperfect. However, a more appropriate measure of capital should take into account the age and depreciation rate of capital stock specific to each firm. This information is not available in our dataset. Nevertheless, we assume that these characteristics are sector-specific and taken into account through these sector-time fixed effects.

both single plant and multi-plant firms are present in our sample. However, productivity is specific to each firm, and not to each plant.

Shocks of productivity and the choice of inputs: the bias of simultaneity

We first estimate specification 2.1 by ordinary least squares (OLS). It is however well-known that production function estimation is plagued by a number of econometric problems. Simultaneity between shocks of productivity and the choice of inputs is certainly the most important problem, as emphasized by Griliches and Mairesse (1995) and, more recently, Ackerberg, Benkard, Berry, and Pakes (2007).[10] Observed inputs (labor and capital) may be correlated with unobserved inputs (managerial ability, quality of land, capacity utilization, etc.) or anticipated productivity shocks. This problem, known at least since Marschak and Andrews (1944), could bias simple OLS estimates.

The IO literature suggests various solutions to cope with that problem. In this paper, we rely on the methodologies developed by Olley and Pakes (1996) (hereafter OP) and Levinsohn and Petrin (2003) (hereafter LP).[11] Olley and Pakes (1996) suggest a consistent semi-parametric estimator for input elasticities. This estimator solves the simultaneity problem by using firm investment decision to proxy unobserved productivity shocks. Broadly speaking, the estimator requires that, for a given level of capital stock, the current level of investment is an increasing function of the unobserved productivity component, so that a higher value of current productivity shocks leads firms to invest more, but that this investment does not affect current stock of capital. A major drawback of this method is that it can only be computed on firms whose investment is strictly positive every year, which drastically reduces the number of observations. In order to mitigate this limitation, Levinsohn and Petrin (2003) suggest using a specific raw material (e.g. electricity), instead of investment, as a strictly increasing function of productivity shocks.

2.3.2 Computing average productivity *per* cluster

The first step provides three measures of productivity (OLS, OP, LP). We draw from each measure an average productivity *per* cluster, i.e. for each area (z)-sector (s)-time (t)

[10]These authors put forward three other sources of errors in the estimation of productivity. 1/ *Endogenous selection:* exits of firms from the market are not exogenous. Ackerberg et al. (2007) note that smaller firms or firms with higher labor/capital ratio are more likely to exit after a negative shock of productivity. Olley and Pakes (1996) also correct for this source of bias. 2/ *Measurement errors in output and inputs:* the value-added contains not only information on the amount of production but also on the competitive structure of the market in which it operates (see Foster, Haltiwanger, and Syverson, 2008). Datasets often lack information on the quality of inputs (be it labor or capital). Capital is most of the time measured at the book value, which is in fact the way it is recorded in our dataset. So, we can control for labor quality through the share of hours worked by each skill group but the quality of capital is not controlled for. However, we assume that sector-specific part of the age or depreciation rate of the capital stock is taken into account through sector-time fixed effects, which should somewhat reduce capital measurement errors. 3/ *Specification problems:* we assume that the production function is a Cobb-Douglas function in which the value-added is the dependent variable. This specification relies on a strong assumption of separability between the labor and capital inputs from the raw (or intermediate) inputs (see Fuss and McFadden, 1978).

[11]See Ackerberg et al. (2007) for details about alternative solutions.

cell. In the empirical part, the basic areal units are the French employment areas (*zones d'emploi*). Continental France is covered by 341 employment areas, whose boundaries are defined on the basis of daily worker commuting patterns. Broadly speaking, they correspond to local labor markets.

The average firm productivity per cluster is computed as:

$$TFP_{zst} = \frac{1}{N_{zst}} \sum_{i \in (z,s,t)} \hat{u}_{it},$$

with N_{zst} the number of firms per cluster and $\hat{u}_{it}$ the residual from the first-step regression.

In order to compute the average TFP, we have to know the exact location and productivity of each plant. However, in the case of multi-plant firms, we only know productivity at the firm level. Hence, we first compute cluster productivity by restricting our sample to single-plant firms. We check in section 2.6 the robustness of our results to alternative solutions.

The three econometric methods - OLS, OP and LP - do not account for potential correlation between individual inputs and the spatial determinants of productivity. We thus develop a fourth and last measure of cluster productivity by introducing cluster-specific fixed effects (F_{zst}) in the first step. In presence of such a high number of dummies, OP and LP methodologies are difficult to implement, we thus estimate the following specification by OLS:

$$log(VA_{it}) = c_{st} + \theta_{multi} + \alpha_s log(L_{it}) + \beta_s log(K_{it}) + \sum_{q=2}^{3} \delta_{qt} sh_{iqt} + F_{zst} + u_{it}. \quad (2.2)$$

For the same reasons as previously, fixed effects, F_{zst}, can only be defined for single-plant firms, the reason why we estimate specification 2.2 on this restricted sample.

Table 2.1 – Correlation between cluster productivities

	OLS-TFP$_{zst}$	OP-TFP$_{zst}$	LP-TFP$_{zst}$	FE-TFP$_{zst}$
OLS-TFP$_{zst}$	1.00			
OP-TFP$_{zst}$	0.83	1.00		
LP-TFP$_{zst}$	0.92	0.80	1.00	
FE-TFP$_{zst}$	0.96	0.79	0.88	1.00

Notes: (i) OLS-TFP$_{zst}$ is the average productivity of a firm located in the cluster zst when computed by Ordinary Least Squares - OP, LP and FE stand respectively for Olley-Pakes method, Levinsohn-Petrin method and Fixed Effect estimation. (ii) All variables are in logarithm. The number of observations stands at 242 178. (iii) All correlations are significant at 1%.

None of our estimation strategies correct for the same bias. The consistency of our results across econometric strategies can be taken as a proof for the robustness of our findings. Note first that average cluster productivities are strongly correlated across methods, as suggested by table 2.1. Second, table 2.2 suggests that these productivities have almost the same distribution. Note also in table 2.2 that the number of observations is smaller for Olley-Pakes and Levinsohn-Petrin methods.

Table 2.2 – Summary statistics for cluster productivity

	# clusters	Mean	St. Dev.	Min	Q25	Q50	Q75	Max
OLS-TFP$_{zst}$	311,698	-0.04	0.35	-6.90	-0.17	-0.03	0.11	5.32
OP-TFP$_{zst}$	242,209	-0.04	0.34	-7.78	-0.18	-0.03	0.11	5.13
LP-TFP$_{zst}$	311,643	-0.05	0.38	-7.28	-0.20	-0.04	0.12	4.78
FE-TFP$_{zst}$	311,698	0.00	0.36	-7.35	-0.13	0.01	0.15	14.25

Notes: (i) The theoretical number of clusters is 181 industries$\times$341 employment areas$\times$11 years=678,931. (ii) All variable in logarithm. St. Dev.=Standard Deviation. Q25, Q50, Q75 are 25th, 50th and 75th percentiles respectively.

2.3.3 Second step: explaining disparities in average firm productivity across clusters

The second step regression consists in explaining disparities in average productivity across clusters by various agglomeration variables.

$$TFP_{zst} = \alpha_{st} + URB_{zt}.\beta + LOC_{zst}.\gamma + X_{zt}.\rho + \mu_{zst}, \qquad (2.3)$$

where TFP_{zst} is the cluster productivity (in logarithm). URB_{zt} are proxies for urbanization economies, namely the size of the market, its accessibility and the diversity of its economic activity. LOC_{zst} refers to proxies for localization economies: the degree of local specialization, the quality of the labor force in the cluster, and the degree of local competition. α_{st} are sector-specific fixed effects that control for the fact that high productivity sectors may have a propensity to locate in specific areas. In other words, we compare the average productivity of firms operating in the same sector, but located in different areas of France.

In this second step, we use weighted least square (WLS) where the weights are the number of plants by cluster. It allows us to give more weight to clusters where average productivity is more accurately estimated and make this second-step regression more in line with the individual level approach.

Note that parameters β and γ are the same for all sectors. According to this assumption, urbanization and localization economies have the same magnitude in each sector. This strong assumption is relaxed in the last section where we allow these elasticities to be sector-specific.

Simultaneity bias

This second-step regression suffers from a number of pitfalls. As emphasized by Combes et al. (forthcoming), agglomeration and productivity may be simultaneously determined. Some areas benefit from specific features that attract firms and enhance their productivity. In that case, productivity in such locations could be higher even without any production externalities. Some good proxies for these features are difficult to find. We introduce in our second-step regression five characteristics specific to employment areas that can drive productivity and agglomeration at the same time. These characteristics are:

a dummy for being on a coast line, a dummy for being on a lake, a dummy for being on a mountain, a dummy for the presence of scenic points, a dummy for a motorway access.

An alternative solution is to rely on an instrumental variable approach. But, valid instruments for agglomeration proxies are hard to find. Combes et al. (forthcoming) use geological and historical data to instrument density and market potential (see also Rosenthal and Strange, 2008). They assume that geological instruments are valid ones because they do not impact directly on local productivity, beyond their effect on employment density. Their historical instruments consist in mid-nineteenth century density. Due to the persistence in human location choices, these variables are highly correlated with actual density, but could be fairly considered as exogenous to actual local productivity. Their main conclusion is that the simultaneity bias remains of small magnitude. These authors are able to instrument area-specific determinants of productivity (or urbanization proxies), but do not deal with the problem of area-sector-time determinants, such as specialization.

Selection and spatial sorting

Two further sources of errors have been put forward in the literature: spatial sorting and selection. The first one deals with the spatial sorting of firms according to observed or unobserved determinants of productivity. In this case, the best entrepreneurs are prone to be found in the same places. In return, these places register a higher average productivity although no agglomeration economies are at work. Nocke (2006) provides a theoretical model for such a sorting mechanism. The rich array of individual firm characteristics we use in the first-step regression allows us to partially control for this bias. Combes et al. (forthcoming) suggest in their study on wages that the sorting of workers according to their observed and unobserved skills is an important source of bias. In the empirical part, we are able to control for the quality of the labor force in each plant. However, we are unable to control for sorting on unobservables.

The second source of errors concerns mechanisms of market selection as emphasized by Melitz (2003) and Melitz and Ottaviano (2008). According to their model, in denser market, competition is tougher and thus weaker competitors are more prone to exit. This mechanically leads to a rise in the local average productivity, once again without any agglomeration externalities. The distinction between spatial sorting and selection is subtle. Okubo et al. (2008) develop a New Economic Geography-type model where firms do not exit from the market in order to escape competition, but relocate. They also conclude to the spatial sorting of firms according to their productivity.

Distinguishing between agglomeration and selection/sorting mechanisms is an hard task, beyond the scope of this paper (see Combes et al., 2009). However, we test the robustness of our findings to the introduction of region-specific dummies in the second step. Agglomeration economies are thus estimated by comparing firms in different employment areas of the same region. The implicit assumption is to consider that selection and/or sorting takes place at the regional level, making firms in the same region the good comparison point.

2.4 Proxies for urbanization and localization economies

2.4.1 Urbanization economies

Proxies for urbanization economies are computed for each year and each French employment area. We only consider the 341 employment areas mapping continental France and exclude Corsica[12] from the analysis. These proxies are computed using employment information for over 10 million plants.

The size of the market: the local density of employment

The first important question is to know whether or not productivity is higher in locations where economic activity is more agglomerated. According to various mechanisms highlighted in the introduction, the extent of external scale economies can be limited by the extent of the market. Conversely, agglomeration economies should increase with the size of the local market. We proxy the local market size by the density of total employment, defined for an employment area z at time t by:

$$density_{zt} = \frac{emp_{zt}}{surf_z}, \tag{2.4}$$

where emp_{zt} is the level of employment in area z (number of full-time workers) at time t and $surf_z$ is the surface area.

Table 2.3 provides summary statistics concerning urbanization and localization proxies for the year 2004. In 2004, the average employment density stands at 60 full-time workers per square kilometer (to be compared with the population density for France, around 112 residents per square kilometer). Employment density distribution is highly skewed. Half of the employment areas register a density with less than 12.5 full-time workers per square kilometer.

As mentioned earlier, employment areas are defined according to daily commuting patterns. By way of consequences, their surface areas are also disparate. Unsurprisingly, the densest employment areas are the smallest ones. The correlation between employment density and surface area stands at -0.62. Then, for a given employment density, differences in the spatial extent of local market can be large, the reason why we introduce, next to employment density, surface area as control.

Market access and *between-area* interactions

Employment areas are not isolated islands, but they form a large contiguous space. Not only may firms benefit from the access to large input and output markets in their area of location, but they may also take advantage of the markets in the neighboring areas. Interactions could spill over the employment area boundaries, leading to the existence of *between-area* interactions. A common proxy for these *between-area externalities* is the so-called *market potential*. In line with New Economic Geography, Head and Mayer (2004)

[12] Corsica is an island. Thus, location choices on this island do not react to the same forces as on the continental territory.

Table 2.3 – Summary statistics for agglomeration proxies

	Mean	St. Dev.	Min	Q25	Q50	Q75	Max
Employment (# workers)	31,675.43	57,021.77	1,892	9,047	16,895	33,174	693,284
# plants	3,069.65	5,243.49	234	1,041	1,721	3,081	72,706
Urbanization Economies							
Density (worker/km^2)	59.39	382.80	0.94	6.91	12.09	24.38	6,577.65
Surface Area (km^2)	1,569.75	986.68	44.93	837.51	1,420.86	2,066.54	6,207.74
Market Potential	119.83	192.81	34.06	53.72	67.83	103.83	1,948.07
Diversity (3-digit level)	29.06	8.59	5.69	23.99	29.99	35.27	51.52
Diversity (2-digit level)	12.40	2.58	3.88	10.95	12.53	14.25	18.74
Localization Economies (3-digit level)							
Employment (# workers)	363.35	1,210.92	0.09	30.37	96.87	303.75	69,256.46
# plants	36.34	131.72	1.00	3.00	9.00	28.00	8,340.00
Specialization	1.79	6.27	0.00	0.44	0.87	1.51	574.52
Share of highly-skilled workers	0.91	0.47	0.00	0.64	0.87	1.10	11.81
Localization Economies (2-digit level)							
Employment (# workers)	985.72	3,433.62	0.10	75.70	264.89	844.20	183,433.80
# plants	96.22	339.79	1.00	6.00	18.00	71.00	14,184.00
Specialization	1.25	2.25	0.00	0.42	0.82	1.33	101.37
Share of highly-skilled workers	0.87	0.38	0.00	0.65	0.85	1.04	5.55

Notes: (i) Summary statistics for the year 2004. (ii) Variables for urbanization externalities are computed across 341 employment areas. (iii) Variables for localization externalities (at the 3-digit level) are computed across 27,943 clusters for the year 2004. Variables for localization externalities (at the 2-digit level) are computed across 10,809 clusters for the year 2004.

and Head and Mayer (2006) (among others) show that market potential could impact on worker wages and firm productivity.

Market potential is computed as the weighted sum of employment density in the neighboring areas, with weights equal to the inverse of distance (between barycenters).

$$Market\ Potential_{zt} = \sum_{z' \neq z} \frac{density_{z't}}{distance_{zz'}}. \tag{2.5}$$

Average market potential stands at 120 for the year 2004, but this variable is also highly skewed. Half of employment areas register a market potential below 67.8, almost half of the mean. Taking into account spatial interaction between areas is all the more important that density is highly spatially autocorrelated. The correlation of employment density and market potential stands at 0.62, suggesting that dense areas are close to each other. This spatial autocorrelation is prone to induce an upward bias in the density elasticity if market potential is not controlled for and externalities spill over boundaries. Indeed, firms in dense areas can more easily benefit from surrounding markets.

Diversity of economic activity

Employment density takes into account the overall size of the market. However, for a given employment size, the distribution of workers across sectors can be very different from one area to the other. Beyond the overall size of the market, urbanization economies can be due to the relative diversity of activities in a given area, as suggested by Jacobs (1969).[13] Such *between-industry* externalities are proxies by an (inverse) Herfindhal index:

$$diversity_{zt} = \frac{1}{\sum_s emp_{zst}^2 / emp_{zt}^2}, \tag{2.6}$$

where emp_{zst} is employment (in full-time workers) in sector s, in area z, at time t. This proxy registers a minimum value equal to 1 if local employment is only concentrated in one sector and increases with the diversity of local economic activity. We consider as sectors items of the 3-digit (NAF220) industrial classification.[14]

2.4.2 Localization economies

Proxies for localization economies are computed for each area-sector-time cluster. In the empirical section, sectors are items of the 3-digit (NAF220) French industrial classification. We consider 181 sectors.[15]

Specialization and *within-industry* externalities

According to Marshallian theories, *within-industry* externalities (or localization externalities) could be of great importance in explaining productivity variations across clusters. For instance, sharing inputs (like labor inputs) or good practices across firms, within both the same sector and the same area could enhance their productivity. Since seminal papers by Glaeser et al. (1992) and Henderson et al. (1995), it is common to proxy these within-industry externalities by an index measuring the relative specialization of area z in sector s. This index is computed by the share of local employment in industry s compared to the same share at the national level.

$$specialization_{zst} = \frac{emp_{zst}/emp_{zt}}{emp_{st}/emp_t}, \tag{2.7}$$

where emp_{zst} is the employment in the cluster zs, emp_{zt} the total employment in area z, emp_{st} the nationwide employment in sector s, and emp_t the nationwide employment at time t.

This index of relative specialization is equal to 1 when the share of local employment in sector s is the same as the overall share of national employment in that sector. When the index stands above 1, area z is relatively specialized in sector s. Note first that the denser areas do not register specific specialization patterns. The correlation between density and specialization stands at -0.13, suggesting that specialized areas are mainly outside dense areas.

Average specialization stands at 1.79 in 2004, suggesting a relative specialization of French employment areas. However, this average value hides a highly skewed distribution

[13] For a theoretical model where the diversity of economic activities in cities drives location choices, see Duranton and Puga (2001).

[14] In section 2.6 we use the 2-digit (NAF60) industrial classification to test for robustness.

[15] In the last section, we test for the robustness of our results to sectoral aggregation.

with a few employment areas registering very high values for the specialization index. Two third of employment areas register a value below 1.5.

Human capital externalities

Skills are unevenly distributed across space as recently emphasized by Combes et al. (2008a). In the first step regression, we control for the quality of the labor force in each firm. We thus control for any sorting effect due to the higher productivity of highly skilled workers. However, local sectoral productivity could still be higher in areas where the amount of skilled workers is large, due to the existence of *human capital externalities*. According to Rauch (1993) and Moretti (2004) (among others), human capital externalities arise if the presence of educated workers makes other workers more productive. These externalities impact on the productivity of firms, beyond the skill composition of their workforce. Hence, the share of skilled workers has to be introduced in the second-step regression to capture these human capital externalities. This index stands between 0 and 12, a further proof of the spatial sorting of workers according to their qualifications or skills.

Competition and local industrial organization

Not only does specialization affect average productivity, but also does the way local sectoral employment is distributed across plants. Rosenthal and Strange (forthcoming) recently suggest that the smallest plants have the biggest effect concerning localization economies. It is the reason why we introduce the total number of plants in the area-sector-time cluster along with the index of relative specialization. The total number of plants is also an easy way to control for local competition on input and output markets.

2.5 Main results

2.5.1 The magnitude of urbanization economies

In our benchmark model, we regress cluster productivity on employment density and surface area, along with the location-specific controls and sector-time dummies. Table 2.4 provides the results. In the bottom panel, some dummies are further introduced in the regression so as to partially control for selection/sorting effects at the regional level.

When regional dummies are absent, the elasticity of average firm productivity to employment density stands between 0.033 and 0.041 depending on the way the first-step productivity is estimated. Once regional dummies are included, this elasticity stands at 0.025 (in the $[0.022 - 0.03]$ range).

These results are in line with Combes et al. (forthcoming). These authors find elasticities ranging from 0.014 to 0.046 depending on the specification and instruments they use. In their setup, individual firm productivity is averaged within employment areas only, but not clusters. Hence, they do not consider the sectoral heterogeneity. Reintroducing this sectoral heterogeneity does not change the average result. Moreover, we do not use an instrumentation strategy. In conclusion, if a simultaneity bias is at work, its magnitude

Table 2.4 – Cluster productivity and employment density

Estimation	OLS-TFP$_{zst}$	OP-TFP$_{zst}$	LP-TFP$_{zst}$	FE-TFP$_{zst}$
Without regional dummies				
Density$_{zt}$	0.033^a	0.035^a	0.041^a	0.036^a
	(0.004)	(0.004)	(0.005)	(0.004)
Surface area$_z$	-0.002	0.0004	0.001	-0.002
	(0.004)	(0.004)	(0.005)	(0.004)
Obs.	311,698	242,209	311,643	311,698
Adj. R^2	0.129	0.127	0.16	0.205
Sector-time dummies	yes	yes	yes	yes
Location-specific controls	yes	yes	yes	yes
With regional dummies				
Density$_{zt}$	0.022^a	0.026^a	0.03^a	0.025^a
	(0.003)	(0.003)	(0.004)	(0.003)
Surface area$_z$	0.011^a	0.014^a	0.018^a	0.013^a
	(0.003)	(0.003)	(0.004)	(0.003)
Obs.	311,698	242,209	311,643	311,698
Adj. R^2	0.151	0.145	0.182	0.225
Sector-time dummies	yes	yes	yes	yes
Regional dummies	yes	yes	yes	yes
Location-specific controls	yes	yes	yes	yes

Notes: (i) Asymptotic robust, clustered (with area-sector blocks) standard error in parenthesis. (ii) ^{a}, ^{b}, ^{c} : Significance at the 1%, 5% and 10% level respectively.

remains small. This point is highlighted in the meta-analysis conducted by Melo et al. (2009).

From an economic point of view, firms located in the densest clusters (i.e. in the 9^{th} decile of the employment density distribution) are, on average, 8% more productive than firms in the least dense areas (i.e. the 1^{st} decile in employment density).[16] This effect is sizeable when compared to the 2.2% annual average productivity growth registered by French firms over 1993-1999 (see Crépon and Duhautois, 2003).

Furthermore, these findings are in line with previous findings in the literature on the subject. In their extensive review, Rosenthal and Strange (2004) report elasticities ranging from 0.03 to 0.11. Results for France are in the bottom part of the range. This finding is not surprising since part of urbanization externalities have already been internalized by French firms.

We argue that comparing results with and without regional dummies provides a credible magnitude range (0.02-0.04) for urbanization economies, even in the presence of selection/sorting effects on unobservables. If location choices are partially driven by unobservable determinants of productivity, the regression without regional dummies will be upward bias. On the contrary, the introduction of regional dummies certainly wipes out part of the

[16] The ratio of employment density at the 9^{th} decile and its value at the 1^{st} decile, the $p90/p10$ ratio, stands at 14.5. Then $exp(0.03 * ln(14.5)) - 1 \approx 0.08$.

productivity premium associated with a location in a denser region. Indeed, part of the productivity premium can be explain by differences in density at the regional level. These differences are captured by the regional dummies. Hence, the regression with regional dummies is certainly downward bias.

Note also that the impact of density on productivity is linear. We try to introduce without any success a square term for density. This variable is not significant, suggesting that French firms could still benefit, on average, from agglomeration.

In table 2.5, we add market potential to the regression. The elasticity to density is almost twice smaller when market potential is introduced. This result is expected due to the high correlation between density and market potential. However, the explanatory power (R^2) of the regression increases, suggesting that market potential impacts on average firm productivity beyond the impact of density. The coefficient of market potential stands between 0.035 and 0.04. The introduction of market potential induces a reduction in the coefficient of density similar to the reduction induced by the introduction of regional dummies. This is not surprising when considering the high correlation between market potential and regional dummies. The R^2 of the regression of market potential on regional dummies stands at 89%. Moreover in the bottom panel of table 2.5 the impact of market potential is drastically reduced when the regional dummies are introduced.

Finally, in table 2.5, we introduce the index of diversity in the regression. This index is not significant in the regression without regional dummies, but registers a small, significant, negative value when regional dummies are introduced. It suggests that diversity has only a minor negative impact on average firm productivity when density is controlled for.

2.5.2 The magnitude of localization economies

So far, we have considered proxies for urbanization economies only. We now introduce our variables for localization economies. The previous results are robust regardless of the measure of productivity. For the sake of simplicity, in the remaining part, we only keep the fixed-effect measure of productivity ($FE - TFP_{zst}$).[17]

In column (b) of table 2.6, we add an index of specialization, while controlling for density, market potential, diversity, sector-time and regional dummies, and location-specific characteristics. The impact of specialization is positive and significant. Its elasticity stands at 0.02. Firms located in areas hosting a relative high share of employment in their industry are, on average, more productive. Note that, contrary to market potential, local specialization has no major impact on the other elasticities, suggesting that its effect is quite orthogonal to urbanization proxies.

The $p90/p10$ ratio for specialization stands at 13.6 (on average across sectors). For a given sector, firms located in areas belonging to the 9^{th} decile for specialization are, on average, 5% more productive than firms located in an area of the first decile for specialization. The impact of specialization is thus less marked than the impact of density but remains important.

[17] We check that results are quantitatively the same with other measures of productivity.

Table 2.5 – The magnitude of urbanization economies

Estimation	OLS-TFP$_{zst}$	OP-TFP$_{zst}$	LP-TFP$_{zst}$	FE-TFP$_{zst}$
Without regional dummies				
Density$_{zt}$	0.022^a	0.025^a	0.029^a	0.025^a
	(0.005)	(0.004)	(0.006)	(0.005)
Surface area$_z$	0.016^a	0.018^a	0.019^a	0.018^a
	(0.004)	(0.004)	(0.005)	(0.005)
Market Potential$_{zt}$	0.037^a	0.035^a	0.038^a	0.038^a
	(0.005)	(0.004)	(0.005)	(0.005)
Diversity$_{zt}$	-0.01	-0.008	-0.005	-0.009
	(0.006)	(0.006)	(0.008)	(0.007)
Obs.	311,698	242,209	311,643	311,698
Adj. R^2	0.143	0.138	0.172	0.217
Sector-time dummies	yes	yes	yes	yes
Location-specific controls	yes	yes	yes	yes
With regional dummies				
Density$_{zt}$	0.02^a	0.024^a	0.03^a	0.023^a
	(0.003)	(0.003)	(0.004)	(0.003)
Surface area$_z$	0.018^a	0.019^a	0.021^a	0.019^a
	(0.003)	(0.003)	(0.004)	(0.003)
Market Potential$_{zt}$	0.019^b	0.016^b	0.006	0.018^b
	(0.008)	(0.008)	(0.011)	(0.008)
Diversity$_{zt}$	-0.013^a	-0.011^b	-0.009	-0.012^b
	(0.005)	(0.005)	(0.006)	(0.005)
Obs.	311,698	242,209	311,643	311,698
Adj. R^2	0.152	0.145	0.182	0.226
Sector-time dummies	yes	yes	yes	yes
Regional dummies	yes	yes	yes	yes
Location-specific controls	yes	yes	yes	yes

Notes: (i) Asymptotic robust, clustered (with area-sector blocks) standard error in parenthesis. (ii) ^{a}, ^{b}, ^{c} : Significance at the 1%, 5% and 10% level respectively.

In column (c) of table 2.6 we introduce the share of highly skilled workers as a further control. Firms located in clusters where the share of highly skilled workers is larger are on average more productive. The elasticity is significant but stands only at 0.007. Moreover, this variable does not add to the explanatory power of the model (R^2), suggesting that its impact on productivity is limited. The introduction of this new control slightly reduces the coefficient of density, due to a positive correlation between these two variables. Hence, even if the existence of human capital externalities can not be denied, their magnitude is rather small, at least beyond the impact of density and specialization.

In the last column of table 2.6, the total number of plants in each cluster is introduced as a new control. We fail to detect any significant impact of this proxy for both competition and the local industrial organization.

Table 2.6 – The magnitude of localization economies

Estimation	(a)	(b)	(c)	(d)
	Urbanization externalities			
Density$_{zt}$	0.023^a	0.024^a	0.023^a	0.022^a
	(0.003)	(0.003)	(0.003)	(0.004)
Surface area$_z$	0.019^a	0.021^a	0.02^a	0.018^a
	(0.003)	(0.003)	(0.003)	(0.004)
Market Potential$_{zt}$	0.018^b	0.012	0.013	0.012
	(0.008)	(0.009)	(0.009)	(0.009)
Diversity$_{zt}$	-0.012^b	-0.012^b	-0.012^b	-0.012^b
	(0.005)	(0.005)	(0.005)	(0.005)
	Localization externalities			
Specialization$_{zst}$		0.021^a	0.021^a	0.02^a
		(0.002)	(0.002)	(0.003)
Sh. of highly-skilled workers$_{zst}$			0.007^a	0.007^a
			(0.002)	(0.002)
# plants$_{zst}$				0.002
				(0.003)
Obs.	311,698	311,698	311,698	311,698
Adj. R^2	0.226	0.233	0.233	0.233
Sector-time dummies	yes	yes	yes	yes
Regional dummies	yes	yes	yes	yes
Location-specific controls	yes	yes	yes	yes

Notes: (i) Asymptotic robust, clustered (with area-sector blocks) standard error in parenthesis. (ii) ^{a}, ^{b}, ^{c} : Significance at the 1%, 5% and 10% level respectively.

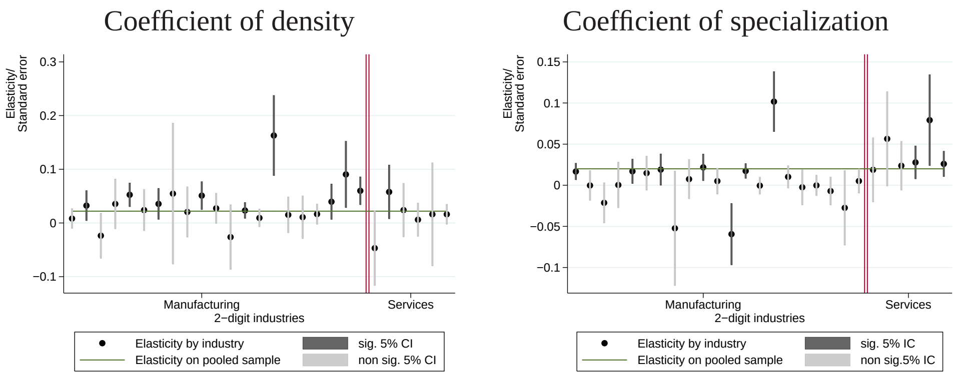
2.5.3 Sectoral heterogeneity

So far, we have considered the average impact of urbanization and localization variables across all sectors. There is no reason why urbanization and localization economies should play with the same magnitude in each sector. In this section, we turn to sector-specific regressions. We focus on employment density and specialization that are the main variables explaining spatial disparities in average productivity.

So as to avoid listing 181 coefficients, we estimate the model of column (d) in table 2.6 for each of the 27 2-digit items of the French industrial classification. Results are reported for density on the left-hand graph of figure 2.1 and for specialization on the right-hand graph of figure 2.1. Three remarks are in order. First, urbanization and localization economies are positive in almost all industries, but significant only in a small subset of industries. It can be partly due to the weak power of our test in the presence of a small number of firms in each industry.

Urbanization effects do not seem to be more prevalent in service industries than in manufacturing ones. On the contrary, localization effects are always positive for service industries, and statistically significant for three (out of six) service industries.

Figure 2.1 – Sectoral heterogeneity



We fail however to find a interesting dimension of heterogenity along which these results could be interpreted. For instance, there is no significant correlation between the spatial concentration of an industry (as measured by an Ellison-Glaeser index for instance) and the magnitude of localization economies.

2.6 Robustness tests

In this section, we test for the robustness of our results along two dimensions: the level of sectoral aggregation and the introduction of multi-plant firms in our sample.

Sensitivity to sectoral aggregation

In previous tables, proxies for urbanization and localization economies are computed at the 3-digit level of the French industrial classification (NAF220). Table 2.7 reports the same results as table 2.6 when urbanization and localization proxies are computed at the 2-digit level (NAF60). Indeed, externalities can work between firms of not only the same 3-digit *sector*, but also the same 2-digit *industry*. Results do not change drastically. Once regional dummies are included, employment density and specialization index remain the first determinants of spatial disparities in average productivity. Productivity elasticity to employment density stands between 0.018 to 0.023, and elasticity to specialization ranges from 0.025 to 0.03. They are of the same order of magnitude than in previous tables. Note that the elasticity to the share of skilled workers is slightly higher in this new setup.

Reintroducing multi-plant firms

In our sample, we only consider single-plant firms with more than 5 full-time workers. Data on production, value-added and capital are unavailable at the plant level. Contrary to Henderson (2003), we are only able to compute firm-level productivity. It is not clear how to redistribute productivity across plants when a firm controls more than one plant. In order to assess the robustness of our results, we develop two polar cases. In the first case, all plants of the same firm are assigned the same firm-level productivity (column *Same productivity for all plants*). In the second case, plant-level productivity is equal to the firm-level productivity cross the share of plant employment in the overall employment of the firm (column *Employment-weighted productivity*).

Results are consistent in both cases (see table 2.8). The employment density and the specialization index have the greatest explanatory power. Elasticity to density is slightly reduced to 0.015, and elasticity to specialization stands between 0.015 and 0.022.

2.7 Conclusion

In this paper, we quantify the magnitude of agglomeration economies on French firm productivity using detailed data from the tax administration. We explain disparities in average firm productivity across clusters by urbanization and localization variables. The large array of individual controls provided in the tax administration data files allows us to purge individual productivity from a number of its determinants unrelated to agglomeration economies, but whose omission could bias our estimates. In particular, we control for the quality of the labor force in each plant and any unobservable sector-specific determinants of productivity.

We employ a two-step procedure. In the first step, we estimate a Cobb-Douglas production function whose residuals is an individual productivity, purged from the aforementionned effects. We then explain disparities in average firm productivity across industrial clusters by, on one hand, employment density, market potential and the diversity of economic structure, and, on the other hand, an index of specialization and the share of local skilled workers. We show that firms located in the densest clusters are, on average, 8% more productive than firms in the least dense areas. This effect is sizeable when compared to the 2.2% annual average growth in productivity registered by French firms over 1993-1999. Not only does local density matters for firms, but also does a good access to surrounding markets. However, we only find a small, negative effect of diversity on productivity, once density is accounted for.

Regarding localization economies, we show that firms located in an area of the 9^{th} decile for specialization are, on average, 5% more productive than firms located in an area of the first decile for specialization. The impact of specialization is thus less marked than the impact of density but remains important. We also find a positive and significant correlation between the quality of the labor force in the cluster and firm productivity, but this variable does not add to the explanatory power of the model. Thus, we cannot deny the existence of human capital externalities. However, once controlled for the quality of the labor input at the plant level, this variable does not impact on productivity beyond the effect of density and specialization.

Recent theoretical developments (see Melitz and Ottaviano, 2008) suggest that firms can be sorted across space according to productivity even within the same sector. This sorting effect is due to a tougher competition in denser markets that force the least productive firm to exit. In order to partially control for such effect, we introduce in our second-step regression regional dummies that control for the average firm productivity at the macro-level. In this setup, the estimation relies on the comparison of average firm productivity across clusters of the same region. Even with this inclusion, our results remain robust.

2.8 Appendix to chapter 2: Data

In this study, we use three different administrative data files, for the years 1994 to 2004. These administrative data files are:

- The SIREN (*Système d'Identification du Répertoire des ENtreprises*) files contain, for each year and in all traded sectors, firm- and plant identifiers (SIREN and NIC code respectively), the municipality code of location of all registered plants, as well as a code (in the 4-digit industrial classification, NAF700) for the main sector of activity (at the plant and firm levels).

- The RSI (*Régime Simplifié d'Imposition*) and BRN (*Bénéfices Régime réel Normal*) files contain the account information declared to the tax administration by each firm in the traded sector. These files provide all the useful information on the output, the value-added (consisting in the output minus the value of intermediary goods), the stock of capital (non-financial assets measured at the book value) at the *firm level*.

- The DADS (*Déclaration Annuelle de Données Sociales*) datasets contain employment information for each plant with at least one paid employee during the year, in the traded and non-traded sectors. This dataset results from the aggregation of individual-level data for each worker paid by the firm. Indeed, the original DADS individual dataset is made upon mandatory employer reports of the gross earnings of each employee subject to French payroll taxes. This file includes around 15 million workers each year. Workers can be followed only through two adjacent years. The files provide information on working days, working hours, wages and various characteristics of the employee (gender, age, occupation) for all plants in the private sector. This file has been collapsed so as to obtain information at the *plant* level and by skill group.

The SIREN dataset contains information for around 3 million plants each year, the DADS file about 1.6 million plants. When matching these two files, we drop plants with zero employees not included in the DADS, and conversely, plants in the finance and real-estate sectors, not registered in the SIREN file. The RSI/BRN file contains about 1.6 to 2 millions firms each year. When merging SIREN/DADS and RSI/BRN files, we drop firms which do not pay taxes (as cooperatives or associations), not included in the RSI/BRN files. The dataset contains about 900,000 firms and 1.2 million plants each year.

A number of selection and correction have been made in order to extract from those raw data files a computationally tractable dataset:

- Aberrant or missing values have been suppressed for value-added, capital and employment. Our unit of observation are not firms *per se*. Indeed, firms can enter or exit the dataset for a number of unknown reasons. As soon as a gap in the spell of observations occurs, we consider that firms before the gap and after the gap are different. Thus, the identifier of a basic unit of observation is both the firm identifier and the starting year of a continuous spell of observations. In the following, we continue, for simplicity, to use the term *firm* but, strictly speaking, it corresponds to a continuous spell of observations for a given firm.

- The main activity of some firms vary across time (even in a given spell of observations), even at the 3-digit classification level. In order to compute sector-specific elasticities, we prefer that firms remain in the same sector across time. We thus consider that the sector of a firm is the one observed during the longest period. If during a continuous spell of observation, the code of activity for a given firm changes more than twice, we drop that spell of observation from the dataset. At this point, we keep 1,762,367 firms, 2,184,811 spells of observation, and 9,186,699 firm-year observations.

Panel of plants

It is important to note that proxies for agglomeration economies, especially density and market potential, are computed on the whole dataset of plants controlled by these 1,762,367 firms. There are more than 10 million plants (10,646,945) in our dataset to compute these agglomeration proxies.

Panel of firms for productivity estimation

In the computation of TFP, we drop firms with strictly less than 5 employees. This is an important choice, as it leads to the deletion of more than 50% observations of the sample each year. However, this selection eliminates noisy data. In addition, it is a quite common selection decision when using those data, see for instance Aubert and Crépon (2003) or Combes et al. (forthcoming).

The Olley-Pakes and Levinsohn-Petrin methodologies require at least two consecutive years of observation, we then drop firms that we observe only one year.

As emphasized in the main text, the second-step regression is computed on the sample of mono-plant firms. Indeed, productivity is computed at the firm-level, and the affectation of productivity to plants in the case of multi-plant firms is more or less arbitrary (see section 2.6. Mono-plants firms account for around 90% of the stock of firms, and 50% of employment an value-added.

In the second-step sample, we only keep 181 sectors from the 3-digit (NAF220) French industrial classification, with at least 100 firms each year. We end with 465,981 firms, corresponding to 3,242,626 firm-year observations.

2.9 Complementary tables

Table 2.7 – Sensitivity to sectoral aggregation

Estimation	(a)	(b)	(c)	(d)
	Urbanization externalities			
Density$_{zt}$	0.023^a	0.023^a	0.018^a	0.011
	(0.003)	(0.003)	(0.003)	(0.007)
Surface area$_z$	0.018^a	0.018^a	0.014^a	0.006
	(0.003)	(0.003)	(0.003)	(0.006)
Market Potential$_{zt}$	0.02^a	0.013	0.018^b	0.015^c
	(0.007)	(0.008)	(0.007)	(0.009)
Diversity$_{zt}$	-0.025^a	-0.022^a	-0.026^a	-0.025^a
	(0.008)	(0.008)	(0.007)	(0.008)
	Localization externalities			
Specialization$_{zkt}$		0.03^a	0.029^a	0.025^a
		(0.004)	(0.004)	(0.003)
Sh. of highly-skilled workers$_{zkt}$			0.041^a	0.042^a
			(0.005)	(0.005)
# plants$_{zkt}$				0.008
				(0.006)
Obs.	119,936	119,936	119,936	119,936
Adj. R^2	0.387	0.403	0.408	0.408
Sector-time dummies	yes	yes	yes	yes
Regional dummies	yes	yes	yes	yes
Location-specific controls	yes	yes	yes	yes

Notes: (i) Asymptotic robust, clustered (with area-sector blocks) standard error in parenthesis. (ii) ^{a}, ^{b}, ^{c} : Significance at the 1%, 5% and 10% level respectively.

Table 2.8 – Sensitivity to the introduction of multi-plant firms

	Same prod. for all plants			Employment-weighted prod.		
Estimation	(a)	(b)	(c)	(d)	(e)	(f)
	Urbanization externalities					
Density$_{zt}$	0.017^a	0.018^a	0.028^a	0.016^a	0.015^a	0.016^a
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Surface area$_z$	0.015^a	0.017^a	0.029^a	0.01^a	0.014^a	0.015^a
	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)	(0.004)
Market Potential$_{zt}$		0.003	0.003		0.013^b	0.01^c
		(0.009)	(0.009)		(0.006)	(0.006)
Diversity$_{zt}$		-0.007	-0.008^c		-0.009^b	-0.009^a
		(0.004)	(0.004)		(0.004)	(0.004)
	Localization externalities					
Specialization$_{zst}$			0.022^a			0.015^a
			(0.003)			(0.003)
Sh. of highly-skilled workers$_{zst}$			-0.00007			0.002
			(0.003)			(0.002)
# plants$_{zst}$			-0.012^a			-0.002
			(0.003)			(0.006)
Obs.	374,956	374,956	374,956	374,956	374,956	374,956
Adj. R^2	0.302	0.302	0.307	0.147	0.147	0.154
Sector-time dummies	yes	yes	yes	yes	yes	yes
Regional dummies	yes	yes	yes	yes	yes	yes
Location-specific controls	yes	yes	yes	yes	yes	yes

Notes: (i) Asymptotic robust, clustered (with area-sector blocks) standard error in parenthesis.
(ii) ^{a}, ^{b}, ^{c} : Significance at the 1%, 5% and 10% level respectively.

CHAPTER 3

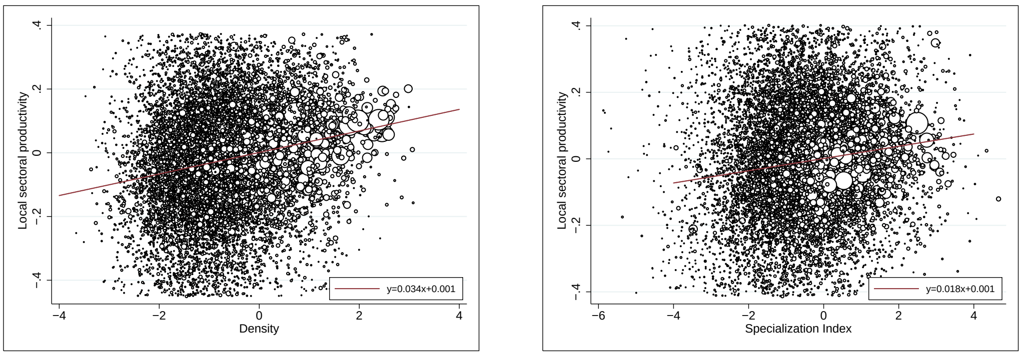
Marshall's scale economies: A quantile regression approach

3.1 Introduction

Firms and workers are, on average, more productive in denser and more specialized areas. This results holds even when we compare firms operating in the same narrowly-defined industry.[1] This is by now a well-established empirical fact, finding support in various countries (see Rosenthal and Strange, 2004; Melo et al., 2009, for surveys). In this paper, we investigate whether this average effect hides large differences across heterogeneous producers.

Figure 3.1 summarizes these results for France, adapted from chapter 2.

Figure 3.1 – Productivity, Density and Specialization



On the left-hand graph, we plot the local sectoral (log) productivity,[2] measured as the average firm (log) productivity within each cluster, against the cluster (log) density of total employment[3]. A cluster is defined as a 3-digit sector in a specific employment area.[4] We

[1]This result holds on average across sectors, but, in chapter 2, we find a large heterogeneity in the magnitude of urbanization and localization economies across sectors. See section 3.6 for further evidence.

[2]In the paper, we consider only productivity in logarithm form. Thus, productivity means log-productivity.

[3]The cluster log density of employment is defined as the logarithm of employment density in the employment area centered around its sectoral mean. See section 3.3.2 for further details.

[4]We use throughout this paper the 3-digit (NES114) French industrial classification to define sectors. Employment areas are spatial units underpinned by clear economic foundations, being defined by the French National Institute of Statistics and Economics (INSEE) so as to minimize daily cross-boundary commuting, or equivalently to maximize the coincidence between residential and working areas.

find an elasticity of average firm productivity to total employment density equal to 0.033.[5] The right-hand graph provides the relationship between local sectoral (log) productivity and local specialization.[6] We find an elasticity to local specialization equal to 0.018. Both results are in line with the literature on the subject, as reported by Rosenthal and Strange (2004).

The size of the dots in these graphs are proportional to the number of firms in each cluster. The local sectoral (log) productivity, i.e. the average (log) productivity across firms within the cluster, summarizes the whole firm (log) productivity distribution in each cluster. Such a traditional approach thus considers that employment density and local specialization impact on all firms in the same way *on average*. The aim of this paper is to question that implicit assumption. We do not only consider that agglomeration economies can impact upon *the average firm productivity* but can also induce some more complex shape shifts in *firm productivity distribution* from one cluster to the other. Extending results from a recent paper by Combes et al. (2009), we claim that heterogeneity among producers can not be disregarded in order to fully understand the impact of agglomeration economies on individual outcomes. To this aim, we use a quantile regression approach to parsimoniously quantify the impact of both urbanization and localization economies at different points in the firm productivity distribution.

The semi-parametric technique of quantile regressions, introduced by Koenker and Bassett (1978) extends the notion of ordinary quantiles to a more general class of linear models in which the conditional quantiles have a linear form. As the linear-in-mean regression model specifies the mean of a conditional distribution as a linear function of a set of regressors, the quantile regression model provides the same parametrization for other moments of the conditional distribution. As recently highlighted by Buchinsky (1998), the quantile regression model has several useful features: 1 - it can be used to characterize the entire conditional distribution of a dependent variable given a set of regressors, 2 - it has a linear programming representation which makes estimation quite easy, 3 - the quantile regression objective function is a weighted sum of absolute deviations so that the estimated coefficients are not sensitive to outliers on the dependent variable, 4 - when the error term is non-normal,[7] quantile regression estimators may be more efficient than least-squares estimators.

Coefficient estimates at distinct quantiles may be interpreted as differences in the response to the changes in the regressors at various points in the conditional distribution of the dependent variable. Namely, in the problem under scrutiny, we can assess the impact of both urbanization and localization economies on firm productivity at different quantiles

[5] We consider in this graph the partial correlation between log productivity and log density once controlled for location-specific characteristics. See section 3.3.1 for details.

[6] Once again, we consider the partial correlation once controlled for the overall size of the local market (through employment density) and its accessibility (through market potential). Local specialization and market potential are defined in section 3.3.2.

[7] This is especially true for models in levels. We only consider in this paper models in logarithm whose distribution of error terms are closer to a normal distribution.

of the conditional productivity distribution.

Quantile regressions have been widely used to study changes in the wage distribution. More specifically, several authors (see Buchinsky, 1998, for a survey) study the returns to education, and their evolution across time, at different points in the wage distribution. As far as we know, quantile regressions have not been used so far to study the impact of urbanization and localization economies on firm productivity.

This paper is related to a recent paper by Combes et al. (2009). Combes et al. (2009) do not only consider the impact of employment density on the average firm productivity but also on the whole distribution of productivity. Relying on trade models with heterogeneous firms *à la* Melitz and Ottaviano (2008), they try to discriminate between two competitive explanations for a higher productivity of firms in denser areas: agglomeration economies or selection effects. In their theoretical model, agglomeration mechanisms induce a right shift of the productivity distribution in denser areas (in comparison with the distribution in sparser areas). By comparison, the selection mechanism induces a left truncation in the distribution. Developing an original quantile approach, they are able to quantify the degree of truncation and right-shifting of the firm productivity distribution in cities with more than 200,000 inhabitants in comparison with cities with less than 200,000 inhabitants. They show that productivity distribution is indeed shifted to the right in denser areas but do not find any evidence of left truncation. They also develop their model one step further and show that not only the productivity distribution is shifted to the right in denser areas, but it is also more skewed to the right. This suggests that agglomeration benefits relatively more the most productive firms. In comparison with Combes et al. (2009), the traditional quantile regression model used in this paper allows us to consider the whole distribution of cluster employment density (not only two subgroups), and to extend the analysis to both urbanization and localization economies. Indeed, there is a long tradition in urban and regional economics to distinguish between these two forms of externalities. In the first case, a firm benefits from the overall size of its market, regardless of the identity of its neighbors. In the second case, a firm benefits from the closeness of neighbors operating in the same industry. Of course, these two categories are not mutually exclusive.

We do find that the impact of employment density, a proxy for urbanization economies, is larger at the right end of the conditional productivity distribution (the 9^{th} decile) than at the left end (the 1^{st} decile). More specifically, quadrupling employment density[8] induces a 3.5% change in productivity for the less productive firms (firms in the first decile of the conditional productivity distribution) against a 6.7% change for the most productive ones (firms in the last decile of the conditional productivity distribution). In other words, the impact of density on productivity is twice larger at the right-end of the firm productivity distribution than at the left-end of that distribution, and the difference is statistically significant. Interestingly, the impact of localization economies is, in contrast, rather stable across the various quantiles of the firm productivity distribution. Quantile elasticities are thus similar to the OLS estimate, at 0.019. It suggests that the traditional linear-in-mean regression

[8]It corresponds to the interquartile ratio in the employment density distribution across clusters.

model does a fairly good job in estimating the magnitude of localization economies but is rather unable to uncover the complex shape shifts in log productivity distribution induced by a shift in total employment density.

These results further question the theoretical literature on agglomeration economies. Indeed, as highlighted by Duranton and Puga (2004), external returns to scale rely on heterogeneity across economic agents. But this heterogeneity is only horizontal, like in the workhorse monopolistic competition model. The previous results need the introduction of some kind of vertical heterogeneity into models to find interpretations. So far, the literature on agglomeration economies and vertical heterogeneity across producers has been quite limited. One exception is Combes et al. (2009) in which workers are more productive when they work for more efficient firms and that this effect is enhanced by interactions with other workers. In their model, initial heterogeneity across producers is magnified by agglomeration economies through the interplay of the productivity of workers. This point certainly deserves further research, especially to understand the difference in results for urbanization and localization economies put forward in this paper.

3.2 Firm TFP estimation: Model and data

Our empirical strategy relies on a two-step approach. In a first step, we compute individual firm productivity controlling for the quality of its labor force in each plant and correcting for simultaneity bias in the choice of inputs by using the Olley and Pakes (1996)'s methodology. In a second step, we quantify how employment density and market potential - proxies for urbanization economies - and local specialization - a proxy for localization economies - impact on the whole distribution of individual firm productivity by using a quantile regression approach. We first present the data at hand before providing more details on the econometric procedure.

3.2.1 Firm and establishment data

Estimating individual productivity requires individual firm and plant information. That information is provided by three different administrative data files, for the years 1994 to 2004.

The SIREN (*Système d'Identification du Répertoire des ENtreprises*) files contain, for each year and in all traded sectors, information about firm- and plant identifiers (SIREN and NIC code respectively), the municipality of plant location, as well as the sector of main activity (in the 4-digit French industrial classification). These pieces of information are available at the plant and firm levels, respectively.

The RSI (*Régime Simplifié d'Imposition*) and BRN (*Bénéfices Régime réel Normal*) files contain the account information declared to the tax administration by each firm in the traded sector. These files provide all the useful information on the output, the value-added (consisting in the output minus the value of intermediary goods), the stock of capital (non-financial assets measured at the book value). This information is provided at the *firm-level* only.

The DADS (*Déclaration Annuelle de Données Sociales*) file contains employment information for each plant with at least one paid employee during the year, in the traded and non-traded sectors. This dataset results from the aggregation of individual-level data for each worker paid by the firm. Indeed, the original DADS individual dataset is made upon mandatory employer reports of the gross earnings of each employee subject to French payroll taxes. This file includes about 15 million workers each year. The file provides information on working days, working hours, wages and various characteristics of the employee (gender, age, occupation) for all plants in the private sector. This file has been collapsed so as to obtain information at the *plant* level and by *skill group*.[9]

The plant-level information (sector, location and hours of work) are used to create proxies for urbanization and localization economies, detailed below. This information is also aggregated at the firm level so as to estimate the first-step production function. Indeed, information about value-added and capital is only known at the firm level. Production function (and thus productivity) can only be estimated at the firm level. We now turn to the estimation of individual productivity.

3.2.2 Production function estimation

We start by constructing the productivity distribution for each location-sector cluster from individual TFP regressions. We consider the 341 French continental employment areas as basic geographical units. We further consider 64 (out of 114)[10] 3-digit items of the French industrial classification as sectors. A cluster is defined as a 3-digit sector in a specific employment area. There is 12,784 clusters (among 21,824 possible clusters) in our sample. Indeed, all sectors are not present in each employment area.

We estimate firm TFP for each sector separately. Note that because information about very small firms tends to be noisy, we only keep firms with more than 5 full-time employees in our sample. Firm TFP is predicted as the residual of a Cobb-Douglas production function

[9]The SIREN dataset contains information for about 3 million plants each year, the DADS file about 1.6 million plants. When matching these two files, we drop plants with zero employee not included in the DADS file, and conversely, plants in the finance and real-estate sectors, not registered in the SIREN file. The RSI/BRN file contains about 1.6 to 2 million firms each year. When merging SIREN/DADS and RSI/BRN files, we drop firms which do not pay taxes (as cooperatives or associations), not included in the RSI/BRN files.

[10]We exclude from our sample banking and insurance because data are unavailable, as well as distribution and consumer services.

estimation:[11]

$$log(VA_{it}) = c_{st} + \theta_{multi} + \alpha_s log(L_{it}) + \beta_s log(K_{it}) + \sum_{q=2}^{3} \delta_{qt} sh_{iqt} + u_{it}, \qquad (3.1)$$

where i indices the firm, and t the year. L_{it} is a measure of employment. In our data, employment is measured as the number of working hours. sh_{iqt} is the share of hours worked by employees of skill group q. Workers are divided up into groups according to their qualification, which aims at controlling for the quality of the labor force. Using the French occupation classification, we set up three categories of skills: (Q3) for highly skilled workers (engineers, technicians and managers), (Q2) for skilled workers (skilled blue and white collars), finally (Q1) for unskilled workers, interns and part-time workers.[12] This specification is justified in Hellerstein, Neumark, and Troske (1999). Combes et al. (2008a) emphasize the spatial sorting of workers according to their qualification in France. Introducing this skill group in equation 3.1 is a simple way to (partially) control for this spatial sorting effect. K_{it} is a measure of the capital stock. In our dataset, this measure consists in the book value of tangible and intangible non-financial assets. Unfortunately we do not have access to details about the quality of capital stock.

We further introduce c_{st} sector-time fixed effects. They control for any sector-time specific determinants of productivity, such as the sector-specific price index for value-added,[13] the sector-specific age and depreciation rate of capital stock,[14] and, finally, any macroeconomic shocks likely to affect value-added and input choices in a specific sector. θ_{multi} is a dummy equal to one if the firm controls more than one plant.

We do not assume constant returns to scale in the production technology. For some industries, the returns appear to be increasing.[15] This wipes out any differences in productivity due to internal returns to scale.

Equation 3.1 is firstly estimated by Ordinary Least Squares (hereafter OLS). It is well known[16] that input elasticities could be biased when estimated by OLS. This is due to either

[11] If agglomeration proxies were introduced in this first step, we would get one elasticity per industry. We thus prefer the two-step approach. Another justification to use a one-step strategy is the introduction of individual fixed effects in the production function. In this case, identification relies on time variation for individual and agglomeration variables. From an empirical point of view, agglomeration variables are rather stable across time and inference in a short panel is dramatically driven by noise and errors in variables. From a theoretical point of view, part of unobservable individual productivity component is also driven by agglomeration externalities, leading to a (potentially strong) downward bias in the estimation. This is the reason why we do not introduce individual fixed effects.

[12] See Burnod and Chenu (2001) for details about this classification.

[13] Indeed, in the production function, the dependent variable is the real value-added of the firm, which is equal to the observed nominal value-added deflated by a sector-time specific price index. This latter term is captured by the fixed effects.

[14] The book-value measure of the stock of capital is imperfect. A more appropriate measure of capital should take into account the age and depreciation rate of capital stock specific to each firm. This information is not available in our dataset.

[15] We do not report the exhaustive list of input elasticities for each industry. However, these coefficients are in line with the previous literature on the subject (see Griliches and Mairesse, 1995).

[16] At least since Marschak and Andrews (1944).

a missing (unobservable) individual determinant of productivity (e.g. quality of managers) or the simultaneity between input choices and productivity shocks. In that latter case, part of the productivity shock, anticipated by the firm manager, but unobservable for the econometrician, drives the choice of inputs. A lot of solutions have been proposed in the literature to cope with that issue (see Ackerberg et al., 2007). In our empirical part, we rely on the strategy developed by Olley and Pakes (1996) (see chapter 2).[17]

As agglomeration economies can take time to materialize in productivity, we choose to compute the average firm productivity over its period of observations as our individual TFP measure. The average period of observations is between 4 and 5 years in our sample. Then,

$$TFP_i = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}, \tag{3.2}$$

where T denotes the number of years the firm is observed, and ε_{it} the residual from equation 3.1.

We are then able to construct not only the average firm productivity but also the whole distribution of individual productivity for each cluster. Note that in this second step, the exact location of each firm has to be known. For multi-plant firms, we know the exact location of each plant but not their productivity. The first step regression only provides a firm-level productivity. This is the reason why we only keep in this second step the 136,474 single-plant firms.[18] Table 3.1 provides basic summary statistics on individual

Table 3.1 – Summary statistics for firm productivity

	# Obs.	Mean	St. Dev.	Q10	Q25	Q50	Q75	Q90	QSC25	QSC10
TFP by OLS	136,474	0.00	0.43	-0.43	-0.21	0.00	0.21	0.46	0.42	0.88
TFP by OP	136,474	0.00	0.45	-0.45	-0.22	0.00	0.23	0.48	0.45	0.94

Notes: (*i*) QSC25 (QSC10) is a quantile-based scale measure at the 25^{th} percentile. (10^{th} percentile respectively.) When the variable is in level, $QSC25 = P75/P25$. When the variable is in logarithm, $QSC25 = P75 - P25$.

firm log productivity when estimated by Ordinary Least Squares (line "TFP by OLS") or by the Olley and Pakes (1996)'s methodology (line "TFP by OP"). These 2 methodologies provide very similar results.

[17]A major drawback of this method is that it can only be computed on firms whose investment is strictly positive every year, which reduces the number of observations. In order to keep as many observations as possible, we use the estimated elasticities to predict the productivity of firms operating in sector s even if the firm is not in the estimation sample.

[18]In chapter 2, we test that this selection does not drastically impact on the results in the linear-in-mean regression setup.

3.3 Agglomeration economies: the traditional linear-in-mean regression model

In this section, we rapidly survey the traditional linear-in-mean regression model and give some benchmark results before turning to quantile regressions.

3.3.1 The traditional linear-in-mean regression model

Basic assumptions

So far, the magnitude of urbanization and localization economies have been estimated using a traditional *linear-in mean regression model*.[19] Given an average cluster measure of productivity, TFP_{zs} (where z indices areas and s sectors), and a set of covariates for urbanization economies (URB_{zs}) and localization economies (LOC_{zs}), the traditional approach specifies the conditional-mean function $\mathbb{E}(TFP_{zs}|URB_{zs}, LOC_{zs}, X_{zs})$ as a linear function of the covariates, while controlling for other determinants of local average firm productivity (X_{zs}), not related to agglomeration economies:

$$\mathbb{E}(TFP_{zs}|URB_{zs}, LOC_{zs}, X_{zs}) = \alpha + URB_{zs}\beta + LOC_{zs}\gamma + X_{zs}\rho. \tag{3.3}$$

In such a model, all regressors have to be centered around their sectoral means.[20] Indeed, in the first-step regression, sector-specific dummies have been introduced to wipe out any sector-specific determinants of productivity. The sector-specific components of urbanization and localization proxies are then to be erased. In other words, the magnitude of agglomeration economies is estimated within the *intra-sectoral* dimension. The identification relies on the comparison between average productivities of firms operating in the same sector but not located in the same area, and thus facing different densities of total employment. Any differences across sectors do not impact on the estimation, as for instance the tendency of a specific sector to locate in denser areas.

The *aggregate*, *cluster-level* model can be estimated by Ordinary Least Squares as followed:

$$\begin{aligned} & TFP_{zs} = \alpha + URB_{zs}\beta + LOC_{zs}\gamma + X_{zs}\rho + u_{zs}, \qquad (3.4)\\ \text{with} \quad & \mathbb{E}(u_{zs}|URB_{zs}, LOC_{zs}, X_{zs}) = 0. \end{aligned}$$

However, from an econometric point of view, model 3.4 is similar[21] to the following

[19] See Combes et al. (forthcoming) and chapter 2 for instance on French data.

[20] An alternative solution is to introduce sector-specific dummies in this second step (see chapter 2). However, such a high number of dummies prevents the quantile regression model from being computed. For the sake of homogeneity between the linear-in-mean and quantile approaches, we prefer to center all regressors around their sectoral means.

[21] For the two models to be perfectly similar, the correct weights have to be introduced in the aggregate model. Namely, the regression has to be estimated by Weighted Least Squares, with weights equal to the number of firms in each cluster zs.

individual-level model:

$$TFP_i = \alpha + URB_{zs}\beta + LOC_{zs}\gamma + X_{zs}\rho + u_i, \tag{3.5}$$
$$\text{with} \qquad \mathbb{E}(u_i|URB_{zs}, LOC_{zs}, X_{zs}) = 0.$$

These models describe how the location of the conditional productivity distribution behaves by only considering the mean of the conditional distribution to represent its central tendency. However, the mean of a distribution provides only a partial information on the way the response variable distribution reacts to a shift in the covariates. The linear-in-mean regression model makes the implicit assumption that the overall distribution is shifted in the same way as its mean when the covariates move.

In the most basic setting, the linear regression model invokes an homoskedasticity assumption, namely that $\mathbb{V}(u_i|URB_{zs}, LOC_{zs}, X_{zs})$, the conditional variance of the disturbance term, is a constant σ^2, independent of i. However, a simple departure from this basic setup allows considering heteroskedasticity, i.e. a shift in the scale of the distribution of the dependent variable when the covariates move. Hence, the traditional linear-in-mean regression model is able to deal with location (mean) and scale (variance) shifts in the shape of the response variable distribution and not more. With our example, the question at hand is to understand whether or not a shift in density (or any other covariates) only induces location and scale shifts in the local sectoral productivity distribution.

Consistency

This second-step regression suffers from a number of pitfalls. As emphasized by Combes et al. (forthcoming), agglomeration and productivity may be simultaneously determined. Some areas may benefit from specific features that attract firms and enhance their productivity. In that case, productivity in such locations could be higher even without any production externalities.

Proxies for such endowments are difficult to find. We introduce in our second-step regression nine location-specific characteristics (corresponding to variables X_{zs} in models 3.4 and 3.5): surface area, longitude, latitude, altitude, declivity, a dummy for being on a coast line, a dummy for being on a lake, a dummy for being on a mountain, contiguity to a national border. By lack of credible instruments, we do not use an instrumental approach as the one proposed by Combes et al. (forthcoming).

Efficiency

Explanatory variables are not firm-specific but area- or area-sector-specific. This mechanically introduces a complex form of correlation between productivity for firms in the same cluster in the *individual-level* model 3.5 (see Moulton (1990), Pepper (2002), Wooldridge (2003)). In the traditional linear-in-mean regression model, correlation patterns can be easily controlled for by an asymptotic robust estimator (with clustering) of the variance. In the quantile regression approach, these correlation patterns are more difficult to deal with. We rely on a block-bootstrap procedure, where blocks are area-sector (zs) clusters.

3.3.2 Proxies for agglomeration economies

Building on the results from chapter 2, we only consider the three following agglomeration economies: employment density and market potential for urbanization economies, and local specialization for localization economies. These proxies are computed for the year 1994,[22] using information from the DADS files for the whole sample (multi- and single-plant firms) in all the 114 original sectors. As highlighted earlier, each variable is centered around its sectoral mean.

The first important question is to know how employment density, a common proxy for the size of the local market, impacts on the distribution of firm productivity. Employment density is defined for a given cluster zs by:

$$ln(Density_{zs}) = ln(Density_z) - \overline{ln(Density_z)}^s$$

with

$$Density_z = \frac{Employment_z}{Surface\, Area_z},$$

where $Employment_z$ is the level of employment in area z (number of full-time workers) and $\overline{ln(Density_z)}^s$ is the average employment density across areas where sector s is located in.

Employment areas are not isolated islands, but they form a large contiguous space. Not only may firms benefit from the access to large input and output markets in the area they are located, but they may also take advantage of the markets in the neighboring areas. Interactions could spill over the employment area boundaries, leading to the existence of *between-area* interactions. A common proxy for these *between-area interactions* is the so-called market potential, computed as the weighted sum of employment density in the neighboring areas, with weights equal to the inverse of distance (between barycenters).

$$ln(Market\, Potential_{zs}) = ln(Market\, Potential_z) - \overline{ln(Market\, Potential_z)}^s$$

with

$$Market\, Potential_z = \sum_{z' \neq z} \frac{Density_{z'}}{distance_{zz'}},$$

where $\overline{Market\, Potential_z}^s$ is the average market potential across areas where sector s is located in. Note that this market potential only takes into account the relative position of employment areas *within* France. However, we know that the integration of European markets can drastically impacts on firm productivity. This is the reason why we introduce different location-specific characteristics that capture the location of each employment area with respect to French borders (longitude, latitude, and contiguity to a national border).

Finally, according to Marshallian theories, *within-industry* externalities (or localization externalities) could be of great importance in explaining TFP variations across clusters. We

[22]The time variation in these variables is very small, so this assumption is benign. Still we do not average proxies for agglomeration economies over time to mitigate any possible reverse causality effect.

introduce an index for the relative specialization of area z in sector s computed as the share of local employment in industry s compared to the same share at the national scale.

$$ln(Specialization_{zs}) = ln(Specialization\,Index_{zs}) - \overline{ln(Specialization\,Index_{zs})}^{s}$$

with

$$Specialization\,Index_{zs} = \frac{Employment_{zs}/Employment_z}{Employment_s/Employment},$$

where $Employment_{zs}$ is the employment in the cluster zs, $Employment_z$ the total employment in area z, $Employment_s$ the nationwide employment in sector s, and $Employment$ the nationwide employment. $\overline{ln(Specialization\,Index_{zs})}^{s}$ is the average local specialization across areas where sector s is located in.

Table 3.3 provides summary statistics for these three variables. For each of them, the first two lines detail moments of the distribution for the variable in level and in logarithm respectively. The third line corresponds to the variable in logarithm once the sectoral mean is subtracted. In that case, all the variables are area- and sector-specific. There are 341 employment areas and 12,784 clusters. Note that employment density and specialization are the two variables with the largest variability. The last two columns of table 3.3 provide two quantile-based scale measures: the interquartile range (QSC25) and the difference between the 1st and 9th deciles (QSC10). For employment density, these statistics stand at 1.29 and 2.88, respectively. This means that a firm located in a cluster with an employment density in the 9th decile faces an almost 18-time as dense environment as a firm located in a cluster with an employment density in the 1st decile[23] than a firm located in a cluster with an employment density in the 1st decile. For specialization, these statistics stand at 1.58 and 3.20, respectively. Similarly, clusters in the 9th decile are almost 25-time as specialized as clusters in the 1st decile.

3.3.3 Results for the traditional linear-in-mean regression model

Table 3.3 provides the results for the traditional linear-in-mean regression model for both measures of productivity - Ordinary Least Squares (OLS) and Olley and Pakes (1996)'s methodology (OP).

Results in both cases are very similar. Columns (A) provide results for the simplest model, where (log) productivity is only explained by the logarithm of total employment density in each cluster. The elasticity of productivity to employment density stands at 0.036, a result similar to the one in the left-hand graph of figure 3.1.[24] The densest clusters (i.e. in the last decile of the employment density distribution) register a 18-time larger density than clusters whose density stands in the first decile. It means that firms located in these latter clusters are, on average, 10.5% ($= 0.036 \times 2.88$)[25] more productive than firms

[23]In this case, we compare density in levels, so $18 = exp(2.88)$.

[24]In figure 3.1, we plot the partial correlation between log productivity and log density once accounted for other spatial determinants of productivity. This is the reason why the results are similar.

[25]We give here a first order magnitude of the productivity premium. The real value stands at: $exp(0.036 \times 2.88) - 1 = 0.109$.

Table 3.2 – Summary statistics for urbanization and localization proxies

Variable	# Obs.	Mean	St. Dev.	Min	Q10	Q25	Q50	Q75	Q90	Max	QSC25	QSC10
						Urbanization proxies						
Density	341	66.12	428.12	0.97	3.87	7.24	12.57	25.76	52.38	7332.26	3.56	13.53
ln(Density)	341	2.71	1.22	-0.03	1.35	1.98	2.53	3.25	3.96	8.90	1.27	2.60
ln(Density)*	12,784	0.00	1.25	-3.24	-1.35	-0.77	-0.16	0.52	1.52	6.19	1.29	2.88
Market Potential	341	118.64	143.22	35.95	47.47	57.21	73.02	118.16	218.32	1244.67	2.07	4.60
ln(Market Potential)	341	4.48	0.65	3.58	3.86	4.05	4.29	4.77	5.39	7.13	0.73	1.53
ln(Market Potential)*	12,784	0.00	0.69	-1.16	-0.66	-0.46	-0.20	0.30	0.94	2.70	0.76	1.59
						Localization proxies						
Specialization	12,784	1.72	5.33	0.00	0.12	0.28	0.67	1.57	3.62	347.13	5.57	30.55
ln(Specialization)	12,784	-0.41	1.34	-5.78	-2.13	-1.27	-0.40	0.45	1.29	5.85	1.72	3.42
ln(Specialization)*	12,784	0.00	1.27	-5.27	-1.59	-0.80	-0.01	0.77	1.61	5.41	1.58	3.20

Notes: (*i*) Variables with a star (*) are centered around their sectoral mean. (*ii*) QSC25 (QSC10) is a quantile-based scale measure at the 25^{th} percentile. (10^{th} percentile respectively.) When the variable is in level, $QSC25 = P75/P25$. When the variable is in logarithm, $QSC25 = P75 - P25$.

Table 3.3 – The spatial determinants of productivity:
A traditional OLS approach

	Dependent Variable: Log of firm productivity					
	TFP by OLS			TFP by OP		
	(A)	(B)	(C)	(A)	(B)	(C)
Density	0.036^a	0.031^a	0.032^a	0.038^a	0.032^a	0.033^a
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Market Potential		0.045^a	0.046^a		0.044^a	0.046^a
		(0.005)	(0.005)		(0.005)	(0.005)
Specialization			0.021^a			0.019^a
			(0.002)			(0.002)
# firms	136,474	136,474	136,474	136,474	136,474	136,474
# clusters	12,784	12,784	12,784	12,784	12,784	12,784
Adj. R^2	0.032	0.033	0.036	0.031	0.032	0.034

Notes: (*i*) All variables are in logarithm. All variables are centered around their sectoral mean, and thus cluster-specific. Controls for area-specific endowments are also included (results not shown). (*ii*) Asymptotic robust, clustered (with area-sector blocks) standard errors in brackets. (*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels respectively.

located in the least dense clusters.

The introduction of market potential in columns (B) reduces the coefficient of density to 0.031, which remains a sizable effect. Productivity elasticity to market potential stands at 0.045. However the dispersion of this variable across clusters is less marked than for density (see table 3.2). The QSC10 statistic only stands at 1.59 (in logarithm). The productivity premium for firms in areas of the last decile in comparison with firms in areas of the first decile for market potential stands at 7%.

Finally, in columns (C), we add an index of local specialization along with density and market potential. This proxy for localization economies is unsurprisingly significantly positive, and stands at 0.021. Once more, the economic effect of this variable depends on its variability. The average QSC10 statistic (across all sectors) stands at 3.20, suggesting that, on average, firms in areas of the last decile for specialization are 7% more productive than firms in areas of the first decile.

The traditional linear-in-mean regression model assumes that these effects are, on average, the same for all firms, regardless of their productivity. In other words, the benefits from agglomeration are homogeneous across firms. In the next section, we reconsider this assumption by estimating models (A), (B) and (C) of table 3.3 by a quantile regression approach.

3.4 Agglomeration economies: a quantile regression approach

The traditional linear-in-mean model only considers the impact of agglomeration economies on the mean of the conditional distribution. In this section, we use quantile regressions to assess this impact at different quantiles of the conditional productivity distribution, and not only the mean. We first present the limitations of the linear-in-mean approach, before turning to a more detailed description of the quantile regression techniques. Results are reported in section 3.5.

3.4.1 Limitations of the traditional linear-in-mean regression model

In order to illustrate the limitations of the linear-in-mean regression model, let us consider tables 3.4 and 3.5. Table 3.4 provides moments of firm productivity distribution for the whole sample (first line) and per quartile of cluster density (lines 2 to 5) for both the OLS and OP measures of productivity.

The first line of table 3.4 is similar to the first line of table 3.1 except that individual productivity is explained by sectoral dummies and location-specific controls, in a first step. Lines 2 to 5 provides the same moments for the productivity distribution when the sample is restricted to firms located in clusters belonging to each quartile of the employment density distribution.[26] The first column provides the number of firms. 12459 firms are located in clusters with an employment density in the 1^{st} quartile. This number obviously increases from the 1^{st} to the 4^{th} quartile, from 12,459 to 74,764 firms.

The average productivity of firms located in areas of the first quartile of employment density is of course larger than the average productivity of firms in the least dense clusters. It is shown in the second column of table 3.4. The fourth to first quartile difference in average productivity (line 6, column 2 of table 3.4) stands at 10%. It is obviously of the same order of magnitude as previously. Average firm productivity increases with the logarithm of employment density.

However, in this table, we do not only compare the average firm productivity but also the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of the conditional productivity distribution across clusters belonging to different quartiles of employment density. At the 10^{th} percentile of the conditional log productivity distribution, the density premium (corresponding to a comparison of firms in clusters of the first quartile for employment density to firms in clusters of the last quartile) stands at 5% only (see line 6, column 4 of table 3.4). At the 90^{th} percentile of the conditional log productivity distribution, the density premium stands at 14%, almost three times larger (see line 6, column 8 of table 3.4). It is a first evidence that the impact of density on productivity is not the same for all firms, and is much larger for firms at the right end of the conditional productivity distribution. The larger elasticity to density in the right end than in the left end of the conditional productivity distribution induces an increase in the scale of this distribution as density increases. Proofs are given in the last two columns of table 3.4 through the quantile-based scale measures, QSC25 and QSC10. These statistics increase across density quartiles. Note that the shape shift in the productivity distribution put forward in table 3.4 can not be taken into account by a tradi-

[26] We are interested in the partial correlation between density and productivity, once controlled for sectoral dummies and area-specific controls for endowments. Thus, cluster employment density is firstly regressed on these controls.

Table 3.4 – Summary statistics for the distribution of firm productivity
Per quartile of density

	# Obs.	Mean	St. Dev.	Q10	Q25	Q50	Q75	Q90	QSC25	QSC10
TFP by OLS										
Total	136,474	0.00	0.43	-0.42	-0.20	0.00	0.21	0.45	0.41	0.87
1^{st} quartile	12,459	-0.07	0.41	-0.46	-0.25	-0.06	0.13	0.36	0.38	0.82
2^{nd} quartile	20,109	-0.04	0.40	-0.43	-0.23	-0.03	0.17	0.39	0.39	0.82
3^{rd} quartile	29,142	-0.02	0.39	-0.41	-0.20	-0.01	0.18	0.39	0.38	0.80
4^{th} quartile	74,764	0.03	0.44	-0.41	-0.18	0.03	0.25	0.50	0.43	0.91
Δ 4^{th} qu./$1^{st}qu.$	–	0.10	–	0.05	0.07	0.09	0.12	0.14	0.05	0.09
TFP by OP										
Total	136,474	0.00	0.44	-0.44	-0.22	0.00	0.22	0.48	0.44	0.92
1^{st} quartile	12,459	-0.07	0.42	-0.49	-0.27	-0.07	0.14	0.38	0.41	0.87
2^{nd} quartile	20,109	-0.04	0.41	-0.46	-0.25	-0.04	0.17	0.41	0.41	0.87
3^{rd} quartile	29,142	-0.02	0.41	-0.43	-0.22	-0.02	0.19	0.42	0.41	0.85
4^{th} quartile	74,764	0.03	0.46	-0.43	-0.19	0.03	0.26	0.52	0.45	0.96
Δ 4^{th} qu./1^{st} qu.	–	0.10	–	0.06	0.08	0.10	0.12	0.15	0.04	0.09

Notes: (*i*) We first regress individual productivity and employment density on sectoral dummies and area-specific controls. The reported values are the residuals from that regressions.

tional linear-in-mean regression model. Such a model is only able to deal with mean and (symmetric) scale shifts.

The impact of density is all the more surprising when compared with the impact of specialization. Table 3.5 provides the same kind of information as table 3.4 by quartile of cluster specialization. Once more, we first regress firm productivity and specialization on sectoral dummies, controls for local endowments, employment density and market potential so as to emphasize the partial correlation between productivity and specialization.

Note first that the productivity premium due to local specialization stands at 6% when we consider the average firm productivity (line 6, column 2 of table 3.5), a similar magnitude as the one put forward in the previous section. Contrary to employment density, the productivity premium due to local specialization is rather stable at different points of the productivity distribution, equal to the estimate for the mean at 6 to 8% (see line 6 of table 3.5). It means that an increase in local specialization induces a simple right shift in the productivity distribution, and that the traditional linear-in-mean regression model is well-suited to capture this kind of effects.

3.4.2 The quantile regression model

Quantile regressions are well-suited to analyze the kind of complex shape shifts in distribution put forward in the previous section. The *quantile-regression model*, first introduced by Koenker and Bassett (1978), allows exploring more detailed transformations of the response variable distribution as covariates move.

Similarly to model 3.4, Koenker and Bassett (1978) define the *quantile-regression*

Table 3.5 – Summary statistics for the distribution of firm productivity
Per quartile of specialization

	# Obs.	Mean	St. Dev.	Q10	Q25	Q50	Q75	Q90	QSC25	QSC10
				TFP by OLS						
Total	136,474	0.00	0.42	-0.42	-0.20	0.00	0.21	0.44	0.41	0.86
1^{st} quartile	10,093	-0.05	0.44	-0.49	-0.25	-0.03	0.18	0.42	0.43	0.91
2^{nd} quartile	26,298	-0.01	0.40	-0.42	-0.20	-0.01	0.19	0.41	0.39	0.83
3^{rd} quartile	50,205	-0.01	0.41	-0.41	-0.20	-0.00	0.20	0.42	0.39	0.83
4^{th} quartile	49,878	0.02	0.45	-0.41	-0.18	0.02	0.24	0.49	0.42	0.90
Δ 4^{th} qu./1^{st} qu.	–	0.07	–	0.08	0.07	0.05	0.06	0.07	-0.01	-0.01
				TFP by OP						
Total	136,474	0.00	0.44	-0.44	-0.21	0.00	0.22	0.47	0.43	0.91
1^{st} quartile	10,093	-0.04	0.45	-0.51	-0.26	-0.03	0.20	0.46	0.46	0.97
2^{nd} quartile	26,298	-0.01	0.41	-0.44	-0.22	-0.01	0.20	0.43	0.41	0.87
3^{rd} quartile	50,205	-0.01	0.42	-0.43	-0.21	-0.01	0.21	0.44	0.42	0.88
4^{th} quartile	49,878	0.02	0.47	-0.44	-0.20	0.02	0.25	0.52	0.45	0.96
Δ 4^{th} qu./1^{st} qu.	–	0.06	–	0.07	0.06	0.05	0.05	0.06	-0.01	-0.01

Notes: (*i*) We first regress individual productivity and local specialization on sectoral dummies, area-specific controls, employment density and market potential. The reported values are the residuals from that regressions.

model where the conditional τ^{th} quantile of the outcome is a linear function of the co-variates:

$$Q_\tau(TFP_i|URB_{zs}, LOC_{zs}, X_{zs}) = \alpha(\tau)+URB_{zs}\beta(\tau)+LOC_{zs}\gamma(\tau)+X_{zs}\rho(\tau). \quad (3.6)$$

Consistent estimates for parameters $\alpha(\tau)$, $\beta(\tau)$, $\gamma(\tau)$ and $\rho(\tau)$ can be obtained by estimating the following *firm-level* model:

$$TFP_i = \alpha(\tau) + URB_{zs}\beta(\tau) + LOC_{zs}\gamma(\tau) + X_{zs}\rho(\tau) + u_i, \quad (3.7)$$
$$\text{with} \qquad Q_\tau(u_i|URB_{zs}, LOC_{zs}, X_{zs}) = 0,$$

where $Q_\tau(u_i|URB_{zs}, LOC_{zs}, X_{zs})$ is the conditional τ^{th} quantile of the residual. Parameters $\alpha(\tau)$, $\beta(\tau)$, $\gamma(\tau)$ and $\rho(\tau)$ are specific to the τ^{th} quantile and can be defined for any quantile between 0 and 1, at least theoretically. It is worth emphasizing that contrary to the linear-in-mean regression model there is no *aggregate, cluster-level* model equivalent to model 3.7.

Koenker and Bassett (1978) show that under this linearity assumption, consistent estimators of τ-specific elasticities are obtained by minimizing the *asymmetric absolute loss function* or "check" function:

$$\min_{\alpha\in\mathbb{R},(\beta,\gamma,\rho)\in(\mathbb{R}^{\mathbb{K}})^3} \sum_{i=1}^{N} c_\tau(TFP_i - \alpha + URB_{zs}\beta + LOC_{zs}\gamma + X_{zs}\rho),$$

where

$$c_\tau(u) = (\tau 1[u \geq 0] + (1-\tau)1[u < 0])|u| = (\tau - 1[u < 0])u,$$

Table 3.6 – Employment density and productivity

	Dependent Variable: Log of firm productivity					
	TFP by OLS					
	OLS	(Q10)	(Q25)	(Q50)	(Q75)	(Q90)
Density	0.036^a	0.025^a	0.03^a	0.034^a	0.041^a	0.047^a
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)
Const.	0.000	-0.417^a	-0.198^a	0.003	0.21^a	0.444^a
	(0.002)	(0.005)	(0.002)	(0.002)	(0.003)	(0.006)
	TFP by OP					
Density	0.038^a	0.027^a	0.033^a	0.037^a	0.043^a	0.048^a
	(0.001)	(0.005)	(0.002)	(0.002)	(0.003)	(0.005)
Const.	0.00	-0.442^a	-0.212^a	-0.0003	0.22^a	0.471^a
	(0.002)	(0.005)	(0.002)	(0.002)	(0.004)	(0.006)
Obs.	136,474	136,474	136,474	136,474	136,474	136,474
# cluster	12,784	12,784	12,784	12,784	12,784	12,784

Notes: (*i*) All variables are in logarithm. (*ii*) Bootstrapped, clustered (with area-sector blocks) standard-errors in brackets, 20 replications, 12784 area-sector clusters.(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels respectively.

with $1[\bullet]$ the indicator function. Hence, every individual, firm-level observation is used in the computation procedure for elasticities at each quantile.

The quantile regression model has the main advantage over the linear-in-mean regression model that it makes easy the analysis of the full conditional distribution of the response variable. In the next section, we use this model to quantify productivity elasticities to urbanization and localization economies at different points in the conditional productivity distribution.

3.5 Results for the quantile regression model

3.5.1 Model A

Table 3.6 reports the results from quantile regressions for model (A), for both estimates of productivity (OLS and OP). The 3 quartiles (Q25, Q50 and Q75) are presented in the table as well as the two extreme deciles (Q10) and (Q90). Figure 3.2 plots the results for each decile between 1 and 9. In the table and the figure, we report results for the constant of the regression, called the *typical setting* (see left-hand graph of figure 3.2) and the productivity elasticity to density (see right-hand graph of figure 3.2).

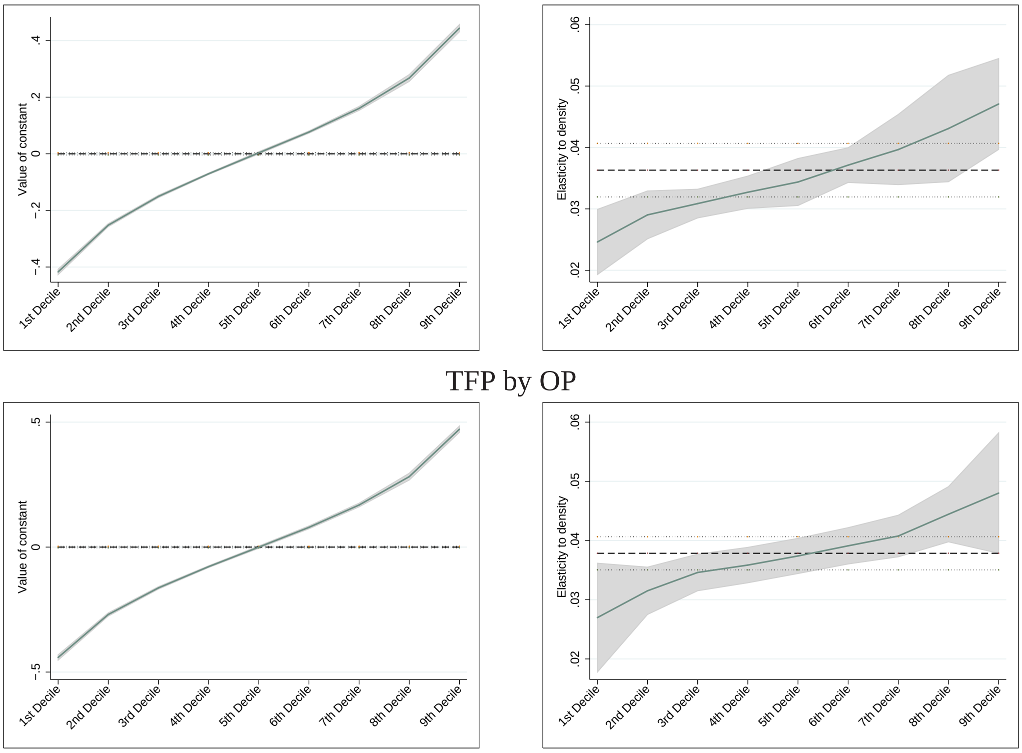
The first column of table 3.6 reports the results for the traditional linear-in-mean regression model. These results are the same as the ones presented in table 3.3 (column A). Note first that all variables in the regression are centered around their sectoral means. It is the reason why the constant term in the OLS regression is equal to 0.

The *typical setting* (left-hand graph of figure 3.2) provides a specific conditional quan-

Figure 3.2 – Employment density and productivity

Typical setting | Elasticity to density per decile

TFP by POLS



tile function for productivity, namely the one fitted at the covariate means. In our case, these means are equal to zero. By definition, the fitted conditional quantile function monotonically increases with the quantiles q. The distribution of productivities lies in the [-0.5, 0.5] range. More interestingly, the median is close to the mean. It is due to the use of a model in logarithm form. The log transformation tends to make the response variable more symmetric around its mean. The steeper slopes at the bottom- and top-end of the graph (1^{st} and 9^{th} deciles) are an evidence of a larger dispersion in productivity at both ends of the distribution.

The effect of employment density (right-hand graph of figure 3.2) can be described as the change in the conditional productivity decile brought about by a shift in cluster employment density. Note first that the density effect is significantly positive at any decile, because the confidence envelope at the 5% level (the shaded area) does not cross the zero line.[27] Second, the right-hand graph of figure 3.2 shows a clear upward-sloping curve for the effect of density. The effect of a shift in employment density is positive for each decile and steadily increasing with deciles. This means that firms at the upper-end of the conditional productivity distribution (the 9^{th} decile) benefit more from a shift in cluster employment density than firms at the lower-end of the distribution (the 1^{st} decile).

[27]This confidence interval is obtained by block-bootstrap.

Table 3.2 provides the QSC10 statistic for cluster employment density, standing at 2.88. Firms in the first decile of the conditional productivity distribution and located in clusters of the 9^{th} decile for employment density are 7.2% (=$exp(0.025 \times 2.88) - 1$) more productive than firms in the same decile of the conditional productivity distribution but located in clusters of the first decile for employment density. When we consider firms in the last decile of the conditional log productivity, the productivity premium due to density increases to 13.5% (=$exp(0.047 \times 2.88) - 1$). The productivity premium is thus twice larger for the most productive firms than for the least productive ones. The difference between the two values is statistically significant. These values are of course in line with the summary statistics provided in table 3.4.

Interestingly, this results is in accordance with Combes et al. (2009), but of a lesser magnitude. They show that cities with more than 200,000 inhabitants are, on average, 10.5 times as dense as cities with less than 200,000 inhabitants. With their estimated elasticities, they find that firms in the bottom decile of the conditional log productivity distribution register a 1% increase in productivity, against a 21% increase for firms in the top decile. The difference can perhaps be explained by the fact that Combes et al. (2009) sort out cities (with various employment densities) into only two groups (below and above 200,000 inhabitants), whereas we consider the whole distribution of cluster employment density. Furthermore, we do not use the same basic geographical units (urban centers in Combes et al. (2009), employment areas in our case). Finally, Combes et al. (2009) only take indirectly into account the spatial industrial structure, i.e. the tendency for specific sectors to locate in high density areas. We highlight in section 3.3 that this effect is controlled for in our setup because proxies for agglomeration economies are centered around their sectoral means. We thus compare differences in productivity for firms operating in the same sector but facing different densities. Due to their methodology, Combes et al. (2009) cannot center the employment density variable. It is the reason why they produce results for each 2-digit industry separatly. Under the assuption that location choices across sectors of the same industry are homogeneous, they also indirectly control for the spatial industrial structure. Table 3.7 provides results for model (A) when we do not center explanatory variables around their sectoral mean. Not only does the elasticity for the conditional mean is lowered from 0.036 to 0.03, but also is the difference between the extreme deciles increased. This result partially fills the gap between our results and the one proposed by Combes et al. (2009).

3.5.2 Model B

Table 3.8 and figure 3.3 provide productivity elasticities to employment density and market potential estimated in model (B). Note first that the introduction of market potential slightly reduces the coefficient of density for all quantiles. The impact of density on productivity remains however statistically significant at all quantiles and increases monotonically with the conditional productivity quantile. The coefficient of density ranges from 0.022 in the 1^{st} decile to 0.043 in the last decile. On the contrary, the impact of market

Table 3.7 – Employment density and productivity
Without centering agglomeration proxies around their sector means

	Dependent Variable: Log of productivity					
	TFP by OLS					
	OLS	(Q10)	(Q25)	(Q50)	(Q75)	(Q90)
Density	0.03^a	0.009^b	0.021^a	0.03^a	0.039^a	0.048^a
	(0.002)	(0.004)	(0.003)	(0.001)	(0.002)	(0.004)
Const.	-0.058	-0.219^c	-0.208^b	-0.031	0.008	0.134
	(0.046)	(0.114)	(0.083)	(0.056)	(0.072)	(0.117)
Obs.	136,474	136,474	136,474	136,474	136,474	136,474
# cluster	12,784	12,784	12,784	12,784	12,784	12,784

Notes: (*i*) All variables in logarithms. (*ii*) Bootstrapped, clustered standard-errors in brackets, 20 replications, 12784 area-sector clusters.(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels respectively.

Table 3.8 – Employment density, market potential and productivity

	Dependent Variable: Log of firm productivity					
	TFP by OLS					
	OLS	(Q10)	(Q25)	(Q50)	(Q75)	(Q90)
Density	0.031^a	0.019^a	0.024^a	0.029^a	0.037^a	0.041^a
	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)	(0.004)
Market Potential	0.045^a	0.047^a	0.044^a	0.041^a	0.039^a	0.045^a
	(0.005)	(0.008)	(0.006)	(0.005)	(0.006)	(0.014)
	TFP by OP					
Density	0.032^a	0.022^a	0.027^a	0.032^a	0.038^a	0.043^a
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.005)
Market Potential	0.044^a	0.046^a	0.045^a	0.043^a	0.038^a	0.038^a
	(0.004)	(0.009)	(0.006)	(0.006)	(0.009)	(0.014)
Obs.	136,474	136,474	136,474	136,474	136,474	136,474
# cluster	12,784	12,784	12,784	12,784	12,784	12,784

Notes: (*i*) All variables in logarithms. (*ii*) Bootstrapped, clustered (with area-sector blocks) standard errors in brackets, 20 replications, 12784 area-sector clusters.(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels respectively.

potential on productivity is almost the same for all deciles. For all deciles, the elasticity to market potential stands around its OLS value, 0.045. This means that market potential impacts upon the conditional productivity distribution through a simple right shift. In that case, we see that the traditional linear-in-mean model does a rather good job in estimating the elasticity to market potential.

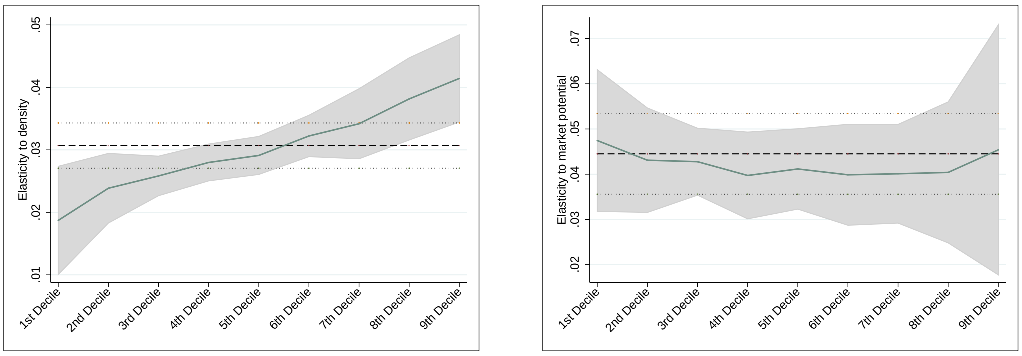
3.5.3 Model C

Table 3.9 and figure 3.4 provide elasticities to density, market potential and local specialization estimated in model (C). The results for density is very similar to the ones obtained in model (B). Contrary to market potential, the introduction of local specialization

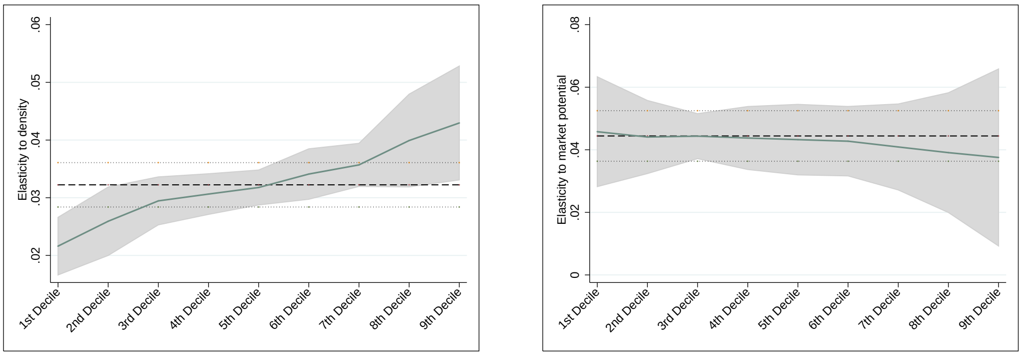
Figure 3.3 – Density, Market Potential and Productivity

Elasticity to density per decile Elasticity to market potential per decile

TFP by POLS



TFP by OP



does not further impact upon the coefficient of density. The impact of specialization on productivity is almost identical to its OLS value for all conditional deciles. However, this impact is a bit larger for extreme deciles, even if the difference is not statistically significant. The conclusion is fairly clear. Contrary to employment density, the impact of specialization is almost uniform across the whole productivity distribution.

Whereas localization economies uniformly shift the productivity of firms, urbanization economies (at least employment density) distort this distribution by a larger increase in productivity at the right end of the distribution than at the left end. This means that firms benefit almost equally from localization economies, but the most productive firms benefit relatively more from urbanization economies than the least productive ones.

3.5.4 Results by industry

In this section, we turn to results by industry. In the second step, we estimate the magnitude of urbanization and localization economies for each 2-digit (NES60) industry separately.[28] We prefer the 2-digit *industry-level* classification rather than the 3-digit *sector-level* classification because any estimation of urbanization and localization economies at

[28]The first-step individual productivity estimation remains computed at the 3-digit sector level.

Table 3.9 – The spatial determinants of productivity: a quantile regression approach

	Dependent Variable: Log of firm productivity					
	TFP by OLS					
	OLS	(Q10)	(Q25)	(Q50)	(Q75)	(Q90)
Density	0.032^a	0.019^a	0.026^a	0.03^a	0.037^a	0.042^a
	(0.001)	(0.004)	(0.003)	(0.002)	(0.002)	(0.004)
Market Potential	0.046^a	0.049^a	0.046^a	0.042^a	0.042^a	0.045^a
	(0.005)	(0.008)	(0.004)	(0.004)	(0.006)	(0.013)
Specialization	0.021^a	0.022^a	0.019^a	0.016^a	0.019^a	0.021^a
	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)	(0.004)
	TFP by OP					
Density	0.033^a	0.023^a	0.028^a	0.033^a	0.038^a	0.044^a
	(0.002)	(0.004)	(0.002)	(0.002)	(0.003)	(0.006)
Market Potential	0.046^a	0.047^a	0.048^a	0.043^a	0.04^a	0.036^a
	(0.004)	(0.008)	(0.006)	(0.005)	(0.007)	(0.013)
Specialization	0.019^a	0.019^a	0.016^a	0.015^a	0.019^a	0.024^a
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.004)
Obs.	136,474	136,474	136,474	136,474	136,474	136,474
# cluster	12,784	12,784	12,784	12,784	12,784	12,784

Notes: (*i*) All variables in logarithms. (*ii*) Bootstrapped, clustered (with area-sector blocks) standard errors in brackets, 20 replications, 12784 area-sector clusters.(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels respectively.

such a detailed level would certainly lead to insignificant results in most cases, due to the small number of observations in each sector. In this subsection model, we only consider (C) where all proxies for agglomeration economies are included.

Table 3.10 in appendix 3.7 provides, for each 2-digit industry, the conditional mean (column OLS) and the typical setting, i.e. the conditional deciles for productivity at the covariate means (equal to zero in our case). The last two columns of table 3.10 provide for each industry the number of firms in the industry and the number of clusters used in the computation of standard errors. Each industry has a zero average productivity. Dispersions in productivity are very similar from one industry to the other, ranging roughly between -0.5 to 0.5. It can be explained by the introduction of sector-specific dummies in the first-step productivity estimation. Such dummies capture all sector-specific determinants of productivity and wipe out any sector-specific components from the residuals. Moreover, we average individual productivity across years. This tends to erase extreme values in productivity.

Table 3.11 in appendix 3.7 provides, for each industry, OLS and quantile estimates for the coefficient of employment density estimated in model (C). Note first that OLS elasticity to density is always positive, but changes a lot from one industry to the other. The impact of density on productivity ranges from 0.011 (significant at the 10% level only) in the electric and electronic equipment industry (E3) to almost 0.092 in the ships, aircraft, railroad equipment industry (E1). Broadly speaking, industries can be sorted out into two types. In the first one, the elasticity to density increases monotonically from the 1^{st} to the 9^{th} decile, with a difference between the two extreme elasticities larger than 0.02 points. This is the case for industry C1, C2, C4, D0, E1, F1, F3, F4 and N2. In these industries, the differential impact of density across deciles is the largest one, suggesting that the most productive firms reap more benefits from the overall size of the market than the least productive firms. For the remaining industries (B0, E2, E3, F2, F5), the elasticity to density is rather stable across deciles, equal to the OLS estimate. For these industries, density impacts upon productivity uniformily across firms. Finally, industry F6 - Electric and Electronic components - has a specific pattern with a larger elasticity to density at the median productivity, but lower values at both ends.

Table 3.12 in appendix 3.7 provides, for each industry, OLS and quantile estimates for the coefficient of market potential estimated in model (C). Note first that OLS estimate is always positive (except for industry C3) but significant (at the 1% level) in a rather limited number of cases (7 out of 16). When significant, the impact of market potential is rather stable across quantiles (see for instance industries E2, E3, F5, N2). In some industries (B0, C1 and D0), the impact of market potential is larger at the lower end of the productivity distribution, suggesting that the least productive firms benefit more from the accessibility to other markets than more productive firms in that specific sectors.

Finally, table 3.13 in appendix 3.7 provides, for each industry, OLS and quantile estimates for the coefficient of local specialization estimated in model (C). As already noted in chapter 2, the elasticity to specialization is positive and significant in a very limited number of industries (6 out of 16). Localization economies are especially strong in industry B0- Food, beverages, and tobacco, E1- Ship, aircraft, railroad equipment, E3 - Electric and Electronic equipment, N2- Consultancy, advertising and business services, C4- Domestic appliances, furniture and finally, E2- Machinery. As noted on the pooled sample, the elasticity to local specialization is rather stable across quantiles. In industry B0 and E3 however, the elasticity to local specialization seems to increase monotonically with quantiles.

3.6 Conclusion

In this paper, we assess the magnitude of urbanization and localization economies on firm productivity, by using a quantile regression approach. The mainstream approach relies on a traditional linear-in-mean OLS approach. So, it is implicitly assumed that agglomeration economies raise the productivity of all firms by the same amount *on average*. The quantile regression approach allows us to question that implicit assumption and to test whether agglomeration economies are not only related to city size but also individual productivity. We are able to test for the differential impact across firms of both urbanization and localization economies.

Two important results stand out from our analysis:

1. Firms are not only more productive in denser areas, but the increase in productivity

induced by urbanization economies[29] is stronger for the most productive firms. This result is true in 9 out of 16 2-digit industries.

2. Firms are more productive in more specialized areas, but localization economies, contrary to urbanization economies, do not benefit more the most productive firms. This result is also true in all industries where localization economies are significantly at work.

These results question the theoretical literature on agglomeration economies. Indeed, as highlighted by Duranton and Puga (2004), external returns to scale rely on heterogeneity across economic agents. But this heterogeneity is only horizontal, like in the workhorse monopolistic competition model. The previous results need the introduction of some kind of vertical heterogeneity into models to find interpretations. So far, the literature on agglomeration economies and vertical heterogeneity across producers has been quite limited. Combes et al. (2009) stands as an exception. They suggest that workers are more productive when they work for more efficient firms and that this effect is enhanced by interactions with other workers. In other words, initial heterogeneity across producers is magnified by agglomeration economies through the interplay of the productivity of workers. There is no doubt that the link between agglomeration economies and vertical heterogeneity deserves further research, especially to understand the differentiated results between urbanization and localization economies put forward in this paper.

[29] Contrary to Combes et al. (2009), our approach does not allow us to disentangle agglomeration economies from selection. However, we take for granted Combes et al. (2009)'s results that there is no significant difference across areas in the intensity of selection (at least at this level of sectoral aggregation).

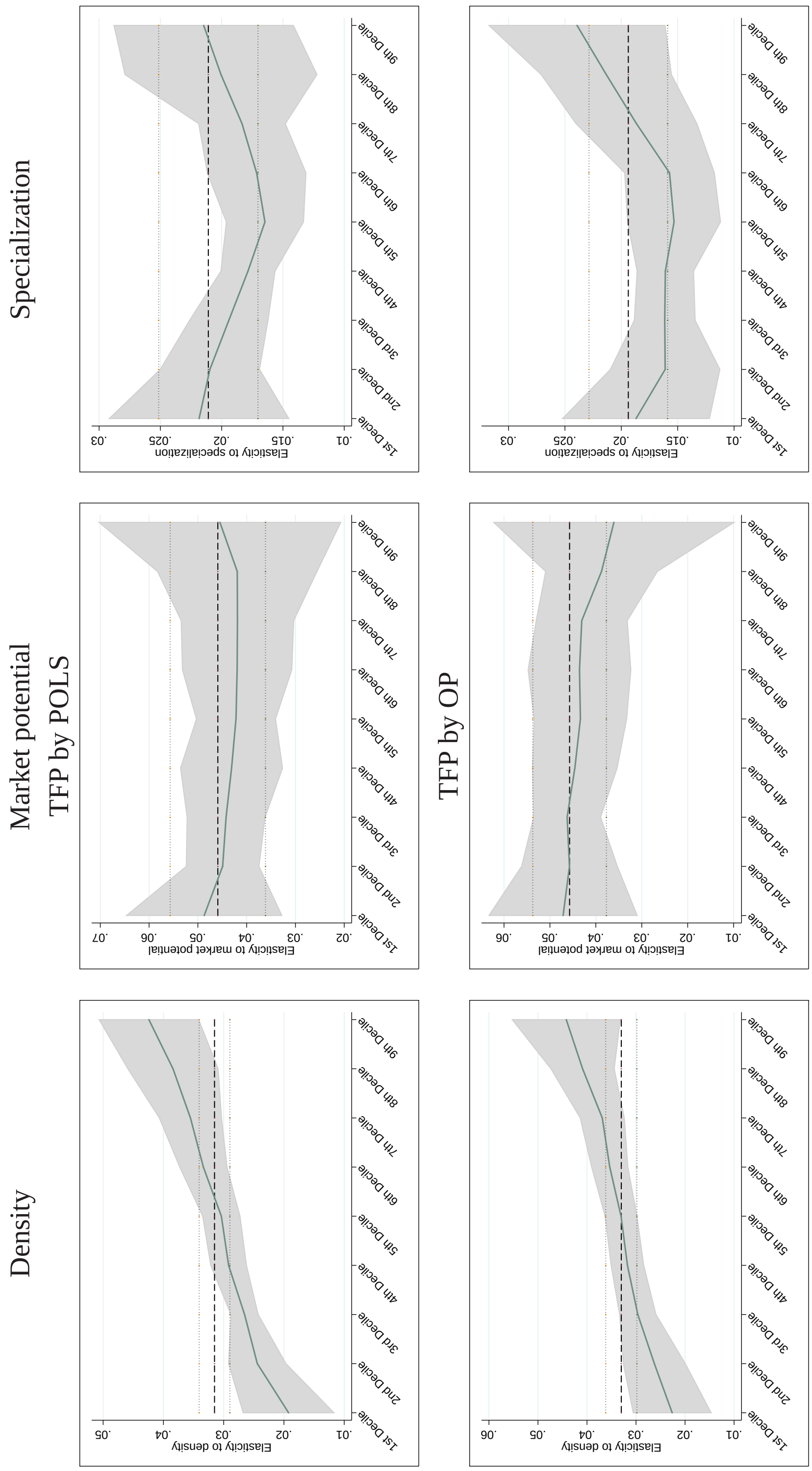


Figure 3.4 – The spatial determinants of productivity

3.7 Appendix to chapter 3: Complementary tables

Table 3.10 – Typical setting: Details by industry

	Dependent Variable: Log of firm productivity							
	TFP by OLS							
	OLS	(Q10)	(Q25)	(Q50)	(Q75)	(Q90)	Obs.	# clust.
B0. Food, beverages and tobacco	0.00	-0.359^a	-0.173^a	0.002	0.176^a	0.373^a	23191	1314
	(0.005)	(0.006)	(0.004)	(0.003)	(0.005)	(0.008)	–	–
C1. Apparel, leather	0.00	-0.455^a	-0.211^a	0.007	0.235^a	0.472^a	4860	466
	(0.011)	(0.029)	(0.021)	(0.011)	(0.011)	(0.017)	–	–
C2. Publishing, printing, recorded media	0.00	-0.43^a	-0.201^a	-0.001	0.206^a	0.454^a	8948	334
	(0.005)	(0.02)	(0.012)	(0.006)	(0.009)	(0.024)	–	–
C3. Pharmaceuticals, perfumes, soap	0.00	-0.57^a	-0.303^a	0.002	0.296^a	0.646^a	804	271
	(0.025)	(0.048)	(0.024)	(0.02)	(0.016)	(0.048)	–	–
C4. Domestic appliances, furniture	0.00	-0.397^a	-0.191^a	0.006	0.207^a	0.425^a	5945	949
	(0.006)	(0.009)	(0.005)	(0.005)	(0.008)	(0.009)	–	–
D0. Motor vehicles	0.00	-0.4^a	-0.167^a	0.008	0.194^a	0.398^a	1378	417
	(0.011)	(0.028)	(0.009)	(0.009)	(0.014)	(0.019)	–	–
E1. Ships, aircraft, railroad equipment	0.00	-0.454^a	-0.192^a	0.022^c	0.227^a	0.482^a	965	271
	(0.022)	(0.023)	(0.014)	(0.012)	(0.016)	(0.043)	–	–
E2. Machinery	0.00	-0.36^a	-0.173^a	0.003	0.184^a	0.384^a	13896	1863
	(0.003)	(0.006)	(0.003)	(0.003)	(0.004)	(0.006)	–	–
E3. Electric and electronic equipment	0.00	-0.369^a	-0.186^a	0.005	0.196^a	0.408^a	5341	950
	(0.005)	(0.01)	(0.005)	(0.007)	(0.006)	(0.014)	–	–
F1. Building materials, glass products	0.00	-0.4^a	-0.195^a	0.002	0.211^a	0.443^a	3967	776
	(0.008)	(0.015)	(0.007)	(0.005)	(0.009)	(0.009)	–	–
F2. Textiles	0.00	-0.431^a	-0.214^a	-0.005	0.224^a	0.47^a	3005	466
	(0.018)	(0.012)	(0.015)	(0.012)	(0.016)	(0.015)	–	–
F3. Wood, paper	0.00	-0.369^a	-0.174^a	0.005	0.19^a	0.397^a	5819	649
	(0.005)	(0.009)	(0.006)	(0.007)	(0.005)	(0.009)	–	–
F4. Chemicals, rubber, plastics	0.00	-0.439^a	-0.201^a	0.008	0.221^a	0.454^a	4696	902
	(0.006)	(0.014)	(0.006)	(0.007)	(0.009)	(0.012)	–	–
F5. Basic metals, metal products	0.00	-0.347^a	-0.169^a	0.003	0.178^a	0.363^a	14085	1254
	(0.006)	(0.009)	(0.003)	(0.006)	(0.006)	(0.007)	–	–
F6. Electric and electronic components	0.00	-0.431^a	-0.186^a	0.012	0.219^a	0.447^a	2244	440
	(0.008)	(0.013)	(0.012)	(0.012)	(0.01)	(0.017)	–	–
N2. Consultancy, advertising, business services	0.00	-0.522^a	-0.247^a	0.006	0.262^a	0.551^a	37330	1462
	(0.006)	(0.007)	(0.006)	(0.005)	(0.008)	(0.017)	–	–

Notes: (*i*) All variables in logarithm. (*ii*) Bootstrapped, clustered (with area-sector blocks) standard errors in brackets, 20 replications.(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels respectively.

Table 3.11 – Elasticity to density: Details by industry

	Dependent Variable: Log of firm productivity							
	TFP by OLS							
	OLS	(Q10)	(Q25)	(Q50)	(Q75)	(Q90)	Obs.	# clust.
B0. Food, beverages and tobacco	0.033^a	0.027^a	0.028^a	0.03^a	0.034^a	0.041^a	23191	1314
	(0.005)	(0.007)	(0.006)	(0.004)	(0.004)	(0.006)	–	–
C1. Apparel, leather	0.023^a	0.004	0.02^b	0.024^b	0.034^a	0.064^a	4860	466
	(0.009)	(0.015)	(0.008)	(0.011)	(0.012)	(0.017)	–	–
C2. Publishing, printing, recorded media	0.038^a	0.015	0.03^a	0.036^a	0.048^a	0.069^a	8948	334
	(0.006)	(0.01)	(0.008)	(0.007)	(0.007)	(0.016)	–	–
C3. Pharmaceuticals, perfumes, soap	0.027	-0.033	0.013	0.034^b	0.058^c	0.059	804	271
	(0.027)	(0.046)	(0.03)	(0.015)	(0.031)	(0.039)	–	–
C4. Domestic appliances, furniture	0.041^a	0.032^b	0.033^a	0.037^a	0.05^a	0.059^a	5945	949
	(0.006)	(0.015)	(0.007)	(0.006)	(0.005)	(0.012)	–	–
D0. Motor vehicles	0.035^a	0.011	0.03^a	0.032^a	0.048^a	0.05^a	1378	417
	(0.013)	(0.029)	(0.01)	(0.012)	(0.012)	(0.013)	–	–
E1. Ships, aircraft, railroad equipment	0.092^a	0.06	0.054^a	0.08^a	0.051^c	0.144^a	965	271
	(0.018)	(0.047)	(0.018)	(0.016)	(0.031)	(0.041)	–	–
E2. Machinery	0.029^a	0.027^a	0.031^a	0.033^a	0.037^a	0.03^a	13896	1863
	(0.004)	(0.007)	(0.003)	(0.003)	(0.004)	(0.006)	–	–
E3. Electric and electronic equipment	0.011^c	0.02^a	0.019^a	0.017^a	0.014^b	0.02^c	5341	950
	(0.007)	(0.007)	(0.004)	(0.004)	(0.006)	(0.011)	–	–
F1. Building materials, glass products	0.021^b	0.01	0.019^c	0.021^b	0.023^b	0.035^b	3967	776
	(0.009)	(0.017)	(0.011)	(0.01)	(0.01)	(0.014)	–	–
F2. Textiles	0.042^a	0.037^c	0.03^a	0.038^a	0.042^a	0.045^b	3005	466
	(0.012)	(0.019)	(0.007)	(0.013)	(0.013)	(0.023)	–	–
F3. Wood, paper	0.045^a	0.011	0.029^a	0.039^a	0.05^a	0.066^a	5819	649
	(0.006)	(0.011)	(0.007)	(0.009)	(0.015)	(0.015)	–	–
F4. Chemicals, rubber, plastics	0.033^a	0.018^b	0.019^b	0.031^a	0.04^a	0.045^c	4696	902
	(0.008)	(0.009)	(0.007)	(0.008)	(0.011)	(0.025)	–	–
F5. Basic metals, metal products	0.023^a	0.019^c	0.02^a	0.022^a	0.023^a	0.024^a	14085	1254
	(0.004)	(0.011)	(0.005)	(0.005)	(0.006)	(0.006)	–	–
F6. Electric and electronic components	0.023^b	0.017	0.022^c	0.036^a	0.031^a	0.0006	2244	440
	(0.011)	(0.02)	(0.011)	(0.01)	(0.011)	(0.027)	–	–
N2. Consultancy, advertising, business services	0.034^a	0.014^b	0.028^a	0.039^a	0.043^a	0.043^a	37330	1462
	(0.004)	(0.007)	(0.008)	(0.004)	(0.005)	(0.01)	–	–

Notes: (*i*) All variables in logarithm. (*ii*) Bootstrapped, clustered (with area-sector blocks) standard errors in brackets, 20 replications.(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels respectively.

Table 3.12 – Elasticity to market potential: Details by industry

	Dependent Variable: Log of firm productivity							
	TFP by OLS							
	OLS	(Q10)	(Q25)	(Q50)	(Q75)	(Q90)	Obs.	# clust.
B0. Food, beverages and tobacco	0.052^{a}	0.071^{a}	0.064^{a}	0.056^{a}	0.048^{a}	0.03^{b}	23191	1314
	(0.013)	(0.016)	(0.014)	(0.009)	(0.011)	(0.014)	–	–
C1. Apparel, leather	0.029	0.129^{a}	0.061^{b}	0.029	-0.005	-0.064	4860	466
	(0.041)	(0.036)	(0.028)	(0.028)	(0.033)	(0.062)	–	–
C2. Publishing, printing, recorded media	0.043^{a}	0.002	0.027	0.051^{a}	0.061^{a}	0.053	8948	334
	(0.015)	(0.022)	(0.017)	(0.018)	(0.013)	(0.043)	–	–
C3. Pharmaceuticals, perfumes, soap	-0.009	0.034	-0.011	0.031	0.049	-0.055	804	271
	(0.071)	(0.116)	(0.072)	(0.058)	(0.06)	(0.129)	–	–
C4. Domestic appliances, furniture	0.031	0.027	0.037^{b}	0.033^{b}	0.021	-0.041	5945	949
	(0.022)	(0.035)	(0.016)	(0.016)	(0.025)	(0.031)	–	–
D0. Motor vehicles	0.059^{a}	0.13^{b}	0.053^{b}	0.057^{c}	0.062^{c}	0.055	1378	417
	(0.022)	(0.055)	(0.025)	(0.031)	(0.032)	(0.05)	–	–
E1. Ships, aircraft, railroad equipment	0.021	0.06	0.079	0.009	0.03	0.109	965	271
	(0.05)	(0.116)	(0.064)	(0.04)	(0.06)	(0.156)	–	–
E2. Machinery	0.045^{a}	0.023^{c}	0.041^{a}	0.039^{a}	0.039^{a}	0.054^{a}	13896	1863
	(0.01)	(0.014)	(0.009)	(0.009)	(0.014)	(0.02)	–	–
E3. Electric and electronic equipment	0.059^{a}	0.054^{b}	0.062^{a}	0.041^{a}	0.033^{c}	0.062^{b}	5341	950
	(0.014)	(0.025)	(0.013)	(0.013)	(0.018)	(0.028)	–	–
F1. Building materials, glass products	0.062^{b}	0.017	0.068^{c}	0.052^{b}	0.05^{b}	0.094^{c}	3967	776
	(0.03)	(0.055)	(0.036)	(0.024)	(0.021)	(0.052)	–	–
F2. Textiles	0.06^{c}	0.074	0.026	0.032	0.035	-0.027	3005	466
	(0.031)	(0.057)	(0.033)	(0.028)	(0.032)	(0.04)	–	–
F3. Wood, paper	0.021	0.043	0.02	0.022	0.023	0.024	5819	649
	(0.026)	(0.028)	(0.025)	(0.014)	(0.024)	(0.043)	–	–
F4. Chemicals, rubber, plastics	0.022	0.043	0.025	-0.009	-0.041	0.019	4696	902
	(0.024)	(0.036)	(0.024)	(0.021)	(0.031)	(0.047)	–	–
F5. Basic metals, metal products	0.046^{a}	0.044^{b}	0.05^{a}	0.051^{a}	0.045^{a}	0.032	14085	1254
	(0.013)	(0.021)	(0.013)	(0.009)	(0.014)	(0.022)	–	–
F6. Electric and electronic components	0.044	0.056	0.044	0.035	0.018	0.033	2244	440
	(0.031)	(0.048)	(0.034)	(0.032)	(0.042)	(0.055)	–	–
N2. Consultancy, advertising, business services	0.056^{a}	0.056^{a}	0.049^{a}	0.051^{a}	0.056^{a}	0.06^{b}	37330	1462
	(0.009)	(0.018)	(0.013)	(0.009)	(0.015)	(0.024)	–	–

Notes: (*i*) All variables in logarithm. (*ii*) Bootstrapped, clustered (with area-sector blocks) standard errors in brackets, 20 replications.(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels respectively.

Table 3.13 – Elasticity to specialization: Details by industry

	Dependent Variable: Log of firm productivity							
	TFP by OLS							
	OLS	(Q10)	(Q25)	(Q50)	(Q75)	(Q90)	Obs.	# clust.
B0. Food, beverages and tobacco	0.066^{a}	0.048^{a}	0.05^{a}	0.053^{a}	0.068^{a}	0.1^{a}	23191	1314
	(0.008)	(0.013)	(0.005)	(0.007)	(0.008)	(0.01)	–	–
C1. Apparel, leather	-0.006	0.019	-0.006	-0.007	-0.0005	-0.012	4860	466
	(0.008)	(0.016)	(0.007)	(0.01)	(0.013)	(0.013)	–	–
C2. Publishing, printing, recorded media	0.013	0.014	0.008	0.006	0.005	0.013	8948	334
	(0.011)	(0.022)	(0.009)	(0.012)	(0.009)	(0.019)	–	–
C3. Pharmaceuticals, perfumes, soap	-0.008	0.026	-0.01	-0.013	-0.027	-0.021	804	271
	(0.019)	(0.027)	(0.019)	(0.023)	(0.028)	(0.037)	–	–
C4. Domestic appliances, furniture	0.026^{a}	0.031^{a}	0.026^{a}	0.025^{a}	0.022^{a}	0.026^{a}	5945	949
	(0.006)	(0.008)	(0.006)	(0.005)	(0.006)	(0.007)	–	–
D0. Motor vehicles	-0.007	-0.019	-0.001	-0.002	-0.014	-0.013	1378	417
	(0.007)	(0.013)	(0.007)	(0.008)	(0.009)	(0.014)	–	–
E1. Ships, aircraft, railroad equipment	0.037^{a}	0.086^{b}	0.032^{a}	0.018^{b}	0.028^{b}	0.053^{a}	965	271
	(0.014)	(0.038)	(0.011)	(0.008)	(0.013)	(0.017)	–	–
E2. Machinery	0.016^{a}	0.018^{a}	0.021^{a}	0.014^{a}	0.015^{b}	0.017^{b}	13896	1863
	(0.003)	(0.005)	(0.005)	(0.003)	(0.007)	(0.007)	–	–
E3. Electric and electronic equipment	0.031^{a}	0.014^{b}	0.025^{a}	0.02^{a}	0.031^{a}	0.032^{a}	5341	950
	(0.008)	(0.007)	(0.004)	(0.004)	(0.006)	(0.006)	–	–
F1. Building materials, glass products	-0.005	0.013	-0.0002	-0.009	-0.02	-0.018^{c}	3967	776
	(0.009)	(0.014)	(0.011)	(0.008)	(0.014)	(0.01)	–	–
F2. Textiles	0.017	0.032^{a}	0.008	0.009	0.007	0.008	3005	466
	(0.011)	(0.012)	(0.009)	(0.01)	(0.009)	(0.011)	–	–
F3. Wood, paper	-0.03^{a}	-0.032^{a}	-0.025^{a}	-0.029^{a}	-0.03^{a}	-0.028^{b}	5819	649
	(0.005)	(0.012)	(0.006)	(0.006)	(0.01)	(0.012)	–	–
F4. Chemicals, rubber, plastics	0.0006	0.003	-0.00006	-0.007	-0.014	-0.008	4696	902
	(0.009)	(0.012)	(0.008)	(0.005)	(0.012)	(0.011)	–	–
F5. Basic metals, metal products	0.006	0.015^{c}	0.006	0.007	-0.0002	-0.005	14085	1254
	(0.005)	(0.008)	(0.005)	(0.005)	(0.005)	(0.007)	–	–
F6. Electric and electronic components	0.002	0.026	0.015	0.009	0.002	-0.006	2244	440
	(0.009)	(0.017)	(0.01)	(0.008)	(0.01)	(0.014)	–	–
N2. Consultancy, advertising, business services	0.026^{a}	0.029^{b}	0.025^{a}	0.022^{a}	0.029^{a}	0.028^{c}	37330	1462
	(0.006)	(0.013)	(0.008)	(0.007)	(0.007)	(0.016)	–	–

Notes: (*i*) All variables in logarithm. (*ii*) Bootstrapped, clustered (with area-sector blocks) standard errors in brackets, 20 replications.(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels respectively.

CHAPTER 4

Product complexity, quality of institutions and the pro-trade effect of immigrants[1]

4.1 Introduction

Despite the widespread availability of modern communication technologies, information costs still play a crucial role in shaping world trade patterns. As surveyed by Anderson and VanWincoop (2004), these costs largely account for the puzzling persistence of distance and border impediments to trade.

According to Rauch (2001), social and business transnational networks are likely to alleviate some of these information failures. Cross-border networks are prone to substitute for organized markets in matching international buyers and sellers, and this is especially true for differentiated products. In this respect, co-ethnic networks are of more particular interest, as illustrated for instance by the model of Casella and Rauch (2003). Immigrants' ties to their home country may promote trade for at least three reasons. First, immigrants have a good knowledge of the customs, language, laws as well as business practices in both the host and home countries. Accordingly, their presence helps bridging the information gap between sellers and buyers on both sides, hence promoting bilateral trade opportunities. Second, immigrant networks may provide contract enforcement through sanctions and exclusions, which substitutes for weak institutional rules and reduces trade costs. In addition to the two previous channels, immigrants bring their taste for homeland products, which should make their trade-creating impact even more salient on imports.

In this paper, we provide new evidence on the relationship between trade and immigration building on regional data for France. We investigate the pro-trade effect of foreign-born French residents on the exports and imports of French *départements* with around 100 countries in the world. The novelty consists in crossing the effect of immigration with both the quality of institutions in the home country and the complexity of traded goods.

The trade-promoting effect of immigration is now well documented (see Wagner et al., 2002, for an extensive review). Gould (1994), Head and Ries (1998) and Girma and Yu (2002) find a significant trade-creating impact of immigrants settled in the United States, Canada, and the United Kingdom respectively. Rauch and Trindade (2002) exhibit a

[1]This paper is joint work with Pierre-Philippe Combes (Univ. of Aix-Marseille & GREQAM) and Miren Lafourcade (Univ. of Paris XI- ADIS & PSE).

diaspora-network rationale ruling this pro-trade phenomenon by showing that South-Asian country pairs with a higher proportion of Chinese immigrants trade more with each other.

However, there are many reasons to suspect that, at the country level, the correlation between trade and immigration might arise from omitted common determinants (such as colonial ties, language or cultural proximity), or reverse causality if immigrants prefer to settle in countries that have good trade relationships with their home country.

Accordingly, a few recent attempts investigate the link between the spatial patterns of trade and immigrants' settlements within countries. Wagner et al. (2002) are the first to test a causal relationship between trade and immigration at the scale of Canadian provinces. The inclusion of country fixed effects allows controlling for the common determinants of trade and immigration at the national level. At the same time, cross-sectional variability in trade and immigration at the regional level provides sufficient information to identify the pro-trade effect of immigrants. The authors confirm the positive and significant elasticity of trade with respect to immigration, at the regional level.

Further evidence is provided for the US state exports. Herander and Saavedra (2005) disentangle the impact of both in-state and out-state stocks of immigrants. The outstanding impact of in-state immigrants pinpoints the key role of local social interactions as a major source of technological externalities. Building on the same previous data set, Dunlevy (2006) further shows that the pro-trade effect of immigrants increases with the degree of corruption and with language similarity in the partner country. Finally, Bandyopadhyay, Coughlin, and Wall (2008) explore the temporal scope of the data and regress the 1990-2000 time variation in trade on the related time variation in immigrant settlements. This approach bears the advantage of controlling for pair-specific unobserved characteristics. The pro-trade effect of immigrants is found to exhibit a large heterogeneity driven by a few countries only. In a related strand of literature, Combes, Lafourcade, and Mayer (2005) for France and Millimet and Osang (2007) for the US show that within-country migrations also affect positively the volume of inter-regional trade flows.

Our paper extends this literature in three directions. First, the relationship between trade and immigration is studied at a lower geographical scale than any previous North-American study. French *départements* are almost 30 times smaller than American states and more than 100 times smaller than Canadian provinces. We do find that immigration exerts a significant positive impact on trade: doubling the number of immigrants settled in a *département* boosts its exports to the home country by 7% and its imports by 4%.

Second, we address econometric questions endemic to gravity-type estimations. We first tackle the issue of specification and selection biases due to zero flows, by using the Quasi-Maximum Likelihood estimator recently proposed by Head, Mayer, and Ries (2009). We then turn to the bias arising from possibly omitted common determinants for immigration and trade or from reverse causality. To circumvent both sources of endogeneity, we include country- and region-specific fixed effects in the regression, and we resort to an instrumental variable approach, where lagged stocks of foreign-born French residents serve as instruments. The previous orders of magnitude remain astonishingly robust to these econometric refinements.

Finally, we evaluate the heterogeneous impact of immigrants on trade along two intertwined dimensions: the complexity of traded goods and the quality of institutions in the partner country. Indeed, Rauch and Trindade (2002) show that the trade-creating effect of Chinese networks is larger for differentiated goods than for homogeneous or reference price goods. The fact that immigrants matter more for differentiated goods can be taken as a support for the information-cost-saving channel of transnational networks. Besides, Anderson and Marcouiller (2002) and Berkowitz et al. (2006) show that the quality of institutions impacts drastically on the volume of bilateral trade. Berkowitz et al. (2006) point out that the quality of institutions matters more for complex commodities, which exhibit characteristics difficult to fully specify in a contract. This is the reason why good institutions may reduce transaction costs when contracts are more incomplete. However, they do not study whether transnational networks could be a substitute for weak institutions, especially in the trade of complex products, as suggested by Rauch (2001).[2]

Building on these insights, we disentangle the pro-trade impact of immigrants across both the partner's institution quality and the complexity of traded goods. In this respect, we emphasize two main results. First, immigrants especially matter for the imports of complex goods, regardless of institution quality in the home country. Turning to the imports of simple products, immigrants matter only when the quality of institutions at home is weak. Second, the trends are less marked for exports. The pro-trade impact of immigrants on exports is positive only when they come from countries with weak institutions, regardless of the complexity of products.

The remainder of the paper proceeds as follows. Section 4.2 presents the augmented-gravity specification we use to evaluate the trade-creating impact of foreign-born French residents, and discusses several econometric issues. It also describes the trade and immigration data for French regions. Section 4.3 presents the benchmark empirical results. Section 4.4 disentangles the trade-creating impact of immigration across simple or complex goods, and across countries with different quality of institutions. Section 4.5 concludes.

4.2 Model specification, econometrics and data

To investigate the pro-trade effect of social networks, we need a benchmark to evaluate the amount of trade expected absent any immigrant settlements. Following Combes et al. (2005), we present the gravity norm we use to provide this benchmark. This section also discusses some econometric pitfalls traditionally encountered in gravity estimations. The following presentation draws on the exposition by Head, Mayer, and Ries (2008).

Model specification

The rationale behind the gravity model is that the value of trade between two locations (y_{ij}) is generated by the adjusted economic sizes of both the supplying location i (S_i) and the demanding location j (M_j), and inhibited by all sources of "trade resistance" between

[2] In this respect, Dunlevy (2006) is a noticeable exception. He shows that the impact of immigrants on US state exports is more important when institutions in the home country are weak.

them (ϕ_{ij}):

$$y_{ij} = GS_iM_j\phi_{ij}, \tag{4.1}$$

where G is a factor that does not vary across regions. Head et al. (2008) refer to S_i and M_j as the monadic terms, and ϕ_{ij} as the dyadic term. The usual practice is to log-linearize this equation and to find proxies for the monadic and dyadic terms:

$$\ln y_{ij} = \ln G + \ln S_i + \ln M_j + \ln \phi_{ij}. \tag{4.2}$$

Anderson and VanWincoop (2003) provide clear-cut theoretical micro-foundations for the monadic terms: they depend on nominal economic sizes (for instance GDP), but also on non-linear functions of all pairwise dyadic terms, called the "Multilateral Resistance Indices". A proper control for these monadic terms in gravity estimations is challenging.[3] The primary question we focus on is whether the spatial distribution of immigrants coming from a country j affects trade flows from hosting *départements* toward that country. Hence, we are not interested in the country- or *département*-specific determinants of trade. This is the reason why we adopt a fixed-effect approach *à la* Anderson and VanWincoop (2003), and introduce two sets of dummies in the gravity equation. The inclusion of country fixed effects (f_j) is meant to control for all standard country-specific determinants of trade: membership to a common trade or currency bloc (e.g. the Euro Zone or the European Union), landlocked nature, colonial ties or common languages. The other set of dummies (f_i) controls for the *département*-specific determinants of trade, such as the density of economic activity or any natural or man-made endowments. Finally, it is worth noting that, in this two-way fixed-effect setting, only the dyadic determinants (ϕ_{ij}) of bilateral trade can be identified.

Regarding this dyadic term, we follow Combes et al. (2005) and assume that trade costs do not only depend on distance and contiguity. They are also inversely correlated with the number of immigrants coming from country j settled in region i. We choose ϕ_{ij} as a multiplicative function of: 1 - the great-circle distance between i and j, 2 - a dummy indicating whether or not the *département* and the country are contiguous,[4] and finally 3 - the stock of foreign-born residents in *départements* i originating from country j, mig_{ij}:

$$\phi_{ij} = \text{dist}_{ij}^{\beta}(1 + \text{mig}_{ij})^{\alpha} \exp(\gamma \text{contig}_{ij}). \tag{4.3}$$

[3]Head et al. (2008) give a clear review of the state-of-art on the econometric specification of the gravity equation. Four solutions are encountered in the literature: 1/ a non-linear approach, proposed by Anderson and VanWincoop (2003), where Multilateral Resistance Indices are explicitly computed, 2/ a fixed-effect approach, also proposed by Anderson and VanWincoop (2003), where monadic terms are controlled for by a set of importer and exporter dummies, 3/ the *bonus vetus OLS* approach, proposed by Baier and Bergstrand (2009) and recently adapted by Behrens, Ertur, and Koch (2007) based on spatial econometrics, where first-order Taylor expansions of Multilateral Resistance Indices are introduced in the specification, and 4/ the *tetrad* approach, proposed by Head et al. (2008), where monadic terms are suppressed thanks to the computation of export ratios.

[4]This dummy is equal to one for only a small subset of *départements* contiguous to Belgium/Luxembourg, Germany, Switzerland, Italy or Spain.

We add an error term (ε_{ij}) that controls for all unobservable dyadic terms uncorrelated with distance, contiguity or the stock of immigrants. The baseline specification we estimate is thus the following two-way fixed-effect log-linearized equation:

$$\ln y_{ij} = f_i + f_j - \beta \ln \text{dist}_{ij} + \gamma \text{contig}_{ij} + \alpha \ln \left(1 + \text{mig}_{ij}\right) + \varepsilon_{ij}. \tag{4.4}$$

In what follows, we estimate this specification for exports and imports separately. We expect parameter β to be negative, and parameters γ and α to be positive.

Econometric issues

Three major econometric problems are usually encountered when estimating gravity models. The first problem deals with the treatment of zero flows. The log-linearized specification (4.4) can only be estimated on strictly positive flows. Various methodologies have been proposed to control for the selection bias arising from keeping positive flows only. Dunlevy (2006) takes the logarithm of one plus the value of the flow as a dependent variable. He also estimates a Tobit model with an arbitrary zero threshold. Herander and Saavedra (2005) use the extended Tobit estimation first proposed by Eaton and Tamura (1994), where the threshold is an ancillary parameter to estimate. This technique, also used by Wagner et al. (2002), rests on a maximum likelihood estimation of the log-linearized model.

A second issue concerns the heteroskedasticity of error terms in levels. In theoretical models, gravity equations take a multiplicative form, as in specification (4.1): hence, if the error term in levels is heteroskedastic, OLS estimates for the log-linearized model are biased.[5] To simultaneously tackle issues arising from zero flows and heteroskedasticity, Santos Silva and Tenreyro (2006) initiated a novel approach by estimating the gravity equation in levels. They propose a easy-to-implement Quasi-Maximum Likelihood (hereafter QML) estimation for the gravity equation, under the assumption that error terms in levels are distributed according to a Poisson distribution. These authors find that the elasticity of trade flows to distance is almost half the magnitude estimated from OLS. However, the Poisson specification builds on the assumption that conditional variance equals conditional mean in the data, $\mathbb{V}(y_{ij}|x_{ij}) = \mathbb{E}(y_{ij}|x_{ij})$. Head et al. (2009) provide a more robust 2-step Negative Binomial (hereafter 2NB) procedure that allows the conditional variance to be a quadratic function of the mean, $\mathbb{V}(y_{ij}|x_{ij}) = \mathbb{E}(y_{ij}|x_{ij}) + \eta^2 \mathbb{E}(y_{ij}|x_{ij})^2$.[6] Hence, in what follows, we compare baseline OLS and 2NB estimates in order to test whether the pro-trade effect of immigrants is robust to these two presumably important biases: zero flows and heteroskedasticity in levels.

[5] This is due to Jensen's inequality, according to which the expected value of the logarithm of a random variable is not equal to the logarithm of the expected value of this variable. Furthermore, the expected value of the logarithm of a random variable depends not only on the expected value of the variable, but also on the other moments of its distribution, especially the variance. Under heteroskedasticity in levels, this variance is a function of explanatory variables, which generates endogeneity in the log-linearized model.

[6] Gourieroux, Monfort, and Trognon (1984) show that QML estimators are consistent as long as the expected value of the dependent variable is well specified, and thus robust to an error in the specification of the true data generating process for the error term. See Cameron and Trivedi (2005) for further details.

The third issue is endogeneity, which may arise from two major sources: omitted variables and reverse causality. At the national scale, one can imagine that preferential links between two countries (resulting from a common colonial history for instance) generate simultaneously trade and immigrant flows. Furthermore, the existence of a strong trade partnership may push people to migrate, creating a reverse causality between trade and immigration. Gould (1994) provides two reasons to believe that cross-section estimations actually preclude the endogeneity bias, at the national level. First, migrations are expected to be more exogenous than trade flows, because they are determined by family reunifications in the first place. As recently analyzed by Thierry (2004), this is also a plausible explanation for France. Second, in addition to family entrance motivations, immigration inflows are conveyed by wage differentials and the pre-existence of a same native/speaking community, rather than by trade opportunities. This is also what suggests the analysis conducted by Bartel (1989) or Munshi (2003) for the US, and by Jayet and Bolle-Ukrayinchuk (2007) for France.

Furthermore, these two sources of endogeneity are partially mitigated when we turn to infra-national data. In specification (4.4), the country- and region-fixed effects control for a large set of common observable and unobservable determinants for trade and immigration flows. Nevertheless, it could be argued that reverse causality and omitted variables are still likely to prevail at the infra-national level. In their study of Canadian province trade flows, Wagner et al. (2002) control, for instance, for the commonality of language, i.e. the probability that a random citizen of a given region speaks the same language as a random citizen of the trading partner. We cannot compute such a variable in the French case. We follow another route and instrument the current stock of immigrants with past stocks in 1975, 1982 and 1990. These lagged stocks are valid instruments as long as they determine the current stock of immigrants, and do not determine current trade flows, beyond their effect on the current stock of immigrants. We provide further support for this view in what follows. The instrumental variable approach has been rarely implemented in the literature.[7]

Data

Trade data consists in exports and imports of the 94 French metropolitan *départements* with around 100 countries. French decentralized customs services record the value of trade flows exclusive of transit shipments, as well as the origin/destination of shipments, i.e. those where goods are actually produced/consumed. Although trade values are available since 1978, we focus exclusively on the recent period to ensure data compatibility with immigrants' stocks. Furthermore, in order to prevent noisy observations due to time-specific shocks (as the euro adoption), we average trade flows over three years (1998, 1999 and 2000) for each *département*-country pairs.

Trade flows are initially available at a very disaggregated industrial level, according to the Standard Goods Classification for Transport Statistics (NST/R classification). We match this classification with the one proposed by Rauch (1999) to characterize the com-

[7] Combes et al. (2005) stands as an exception.

plexity or degree of differentiability of goods.[8]

The 1999 French population census provides us with exhaustive information on the number of foreign-born residents by *département* and country pairs. We define immigrants as residents born abroad with a foreign nationality. We check in an earlier version of this paper that results are quantitatively the same when we consider as immigrants residents born abroad with a French or foreign nationality. In the empirical part, we also use the lagged stocks of immigrants to tackle the endogeneity issue. These figures are provided by French population censuses for the years 1975, 1982 and 1990. Appendix 4.6 provides further details on exports, imports and immigration data.

It is worth stressing that most of the variability in the data comes from the cross-country dimension of the sample. For instance, the regression of trade flows on country-specific dummies returns an adjusted-R^2 of 51% for exports, 61% for imports and 70% for immigration. We wipe out this cross-country variation with a set of country fixed effects. We also include *département* dummies to control for the *département*-level observable or unobservable determinants of trade and immigration flows common across all trading partners.

Due to the introduction of these two sets of dummies, the pro-trade impact of immigrants is identified along the within-country and within-*département* data variability. Table 4.1 depicts the within-country and within-*département* correlation between exports, imports, distance and immigration.[9] As expected, distance is negatively correlated with

Table 4.1 – *Within-country, within-département* correlations

Variables	Exports	Imports	Distance	Immigrants
Exports	1.000			
Imports	0.144	1.000		
Distance	-0.090	-0.137	1.000	
Immigrants	0.066	0.043	-0.090	1.000

Notes: All correlations are significant at the 1% level. Correlations between residuals from the regressions of each variable on the two sets of country-specific and département-specific dummies.

exports and imports, the correlation being stronger for imports. By way of contrast, immigration is significantly and positively correlated with both exports and imports. Distance and immigration are also negatively correlated, as it is well known that immigration flows also share a gravity pattern. Appendix 4.6 provides further summary statistics on the data.

[8]See appendix 4.7 for details.

[9]More formally, this is the correlation between the residuals of the regression of each variable on country-specific and *département*-specific dummies.

4.3 The pro-trade effect of immigrants

4.3.1 Benchmark results

Table 4.2 provides the basic results drawn from estimating specification (4.4). In columns labeled OLS, we report the results drawn from the log-linear form (null flows are left out of the sample). We also estimate the same specification in levels (columns 2NB). We run each specification twice: first on the sample restricted to positive flows (columns (3) and (7)), and second on the whole sample (columns (4) and (8)). We run two sets of regressions, for exports and imports separately.

Table 4.2 – Benchmark results

	Exports				*Imports*			
	In log		In levels		In log		In levels	
	OLS	OLS	2NB > 0	2NB ≥ 0	OLS	OLS	2NB > 0	2NB ≥ 0
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance	-0.81^a	-0.777^a	-0.963^a	-0.961^a	-1.488^a	-1.480^a	-1.612^a	-1.638^a
	(0.089)	(0.085)	(0.1)	(0.104)	(0.128)	(0.127)	(0.143)	(0.157)
Contiguity	0.452^a	0.273^c	0.123	0.099	0.445^b	0.342^c	0.029	-0.0009
	(0.167)	(0.163)	(0.163)	(0.169)	(0.198)	(0.201)	(0.205)	(0.237)
Immigrants		0.102^a	0.091^a	0.109^a		0.054^b	0.094^a	0.089^b
		(0.018)	(0.019)	(0.021)		(0.027)	(0.035)	(0.041)
Obs.	9033	9033	9033	9400	8110	8110	8110	9494
Adj. R^2	0.844	0.844			0.8	0.8		

Notes: Country and *département* fixed effects are not reported here. Robust standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

Log-linear specification

In columns (1) and (5), trade impediments are proxied by distance and contiguity only. Elasticities have expected signs. Exports, as well as imports, decrease with distance and increase with contiguity. The elasticity of exports to distance is half the value for imports. Although there is not any obvious reason for such a phenomenon, it is worth recalling that, in this two-way fixed-effect setting, elasticities are estimated on the within-variability of the data. Hence, identification relies drastically on close countries for which distance differentials across regions remain high in comparison with countries located further away. For instance, Paris and Marseille are almost equally distant from the United States, but not from Germany. For more distant countries, the variability in distance is reduced. Nevertheless, the variability in trade flows remains fairly high: a small difference in distance can be associated with a large difference in trade values.

In columns (2) and (6), we add the stock of immigrants in the specification in logs. Contrary to most of the previous regional studies, we are able to assess separately the impact of immigration on exports and imports. Immigrants have a strongly significant impact. They promote exports as well as imports: doubling their number yields a 7% ($2^{0.102} \approx 1.07$) increase in the value of exports and a 4% ($2^{0.054} \approx 1.04$) increase in the value of imports. The pro-trade effect of immigration on imports is almost half the effect on exports. This casts doubt on the existence of a preference channel. However, we will see later that such a difference, which is barely significant here, is in any case not very robust.

The impact on exports is also almost half the value previously found for U.S. state exports. We argue that previous estimations could be tainted with an upward omitted variable bias that can be controlled for by using country fixed-effects. The impact of distance and contiguity is reduced when the stock of immigrants is accounted for. Contiguity is only significant at the 10% level. Immigrants coming from neighboring countries, such as Belgium, Germany or Italy, locate according to a gravity pattern. Consequently, the share of immigrants originating from these neighboring countries is much higher in the regions near the border than anywhere else in France.

Specification in levels

We push further the evidence by testing the robustness of the results to two kinds of possible biases: specification and selection due to neglecting zero flows in the log-linear specification.

Columns (3)-(4) and (7)-(8) in table 4.2 report the results of the 2-step negative binomial estimation procedure (equation (4.4) in levels). The positive and significant impact of immigrants is confirmed. Furthermore, it is of the same order of magnitude than in the specification in levels: doubling the number of immigrants from a country yields a 6.5% increase in both the values of exports and imports with this trade partner. Hence, the results do not change drastically when moving to a specification in levels. Furthermore, they are not driven by the zero-flow truncation. In columns (4) and (8), where null flows are included in the sample, results remain barely the same.

Finally, we provide further robustness checks based on different estimation techniques (see table 4.11 in appendix 4.8). The orders of magnitude are virtually the same in all procedures but the Poisson QML estimation. This is probably due to the assumption that conditional mean equals conditional variance, which would not be valid in our data. Therefore, the pro-trade effect of immigration is robust to both specification and selection biases. We now turn to the endogeneity problem in the log-linear specification.

4.3.2 An instrumental variable approach

Despite the inclusion of fixed effects and the use of a fine geographical scale, our results could still be plagued by the endogeneity of immigrants' stocks. We use an instrumental variables approach to circumvent this issue within the log-linear model.[10] We choose the lagged stocks of immigrants for the years 1975, 1982 and 1990 as instruments.

Relevance of instruments

[10]Non-linear models, as the negative binomial model, remain quite hard to instrument, as reviewed by Windmeijer (2006). Instrumenting is all the more challenging in our setting that we include numerous dummies. This is the reason why, in this section, we exclusively focus on the log-linear specification.

In order to be relevant, instruments have to be correlated with the current stock of immigrants. Hence, we should observe some persistence in the geography of immigrants' settlements within France, by country of origin. This is a well-known empirical fact. For instance, Jayet and Bolle-Ukrayinchuk (2007) find that, in France, past settlements strongly determine the location of new immigrants, due to the existence of social networks or to family motives. Table 4.3 reports the pairwise correlations between past and current stocks of immigrants. We see that these correlations are indeed fairly high, even though they decrease as time-lag raises. This is a first support for validating instruments.

Nevertheless, strict relevance depends on the partial correlation between the endogenous variable and the instruments, once the other exogenous regressors have been controlled for. Table 4.4 reports the OLS estimates of the traditional first step of the 2-step instrumented regression. We further report the F-test of the joint significance of excluded instruments, as well as the Bound, Jaeger, and Baker (1995) partial R^2 (BJB R^2 herafter). As shown by Baum, Schaffer, and Stillman (2003), in the case of a single endogenous explanatory variable, these tests are sufficient to assess the relevance of instruments. According to the Staiger and Stock (1997) rule of thumb,[11] our instruments are relevant. Nevertheless, in regression (4), the elasticity of the 1968 stock of immigrants is not significant. The weakness of instruments being often worse that the endogeneity bias itself, we choose to remain parsimonious, and leave this instrument out of the list.

Table 4.3 – Pairwise correlations for instruments

	ln(1+Immigrants 1999)	
	Correlation	Nb. obs.
ln(1+Immigrants 1990)	0.92	8011
ln(1+Immigrants 1982)	0.92	5697
ln(1+Immigrants 1975)	0.87	4366
ln(1+Immigrants 1968)	0.79	4162

Note: All correlations are significant at the 1% level.

Supporting the validity of instruments

In what follows, we estimate two instrumented models. In the first one, we use the stock of immigrants in 1990 as the only instrument. This variable is actually the most highly correlated with the endogenous regressor, and it is non-missing for most of the observations. Consequently, the model is just-identified and the validity of the instrument, which cannot be tested, must be assumed. In the second model, we run a GMM-type instrumentation by introducing simultaneously the lagged stocks of immigrants in 1975, 1982 and 1990. Even though the number of missing observations drastically increases, the model is now over-identified. Hence, we can test for over-identification restrictions. We follow the suggestion of Baum et al. (2003) in the presence of heteroskedasticity, and run the Hansen-J test. A rejection of the null hypothesis implies that the instruments do not

[11] In the case of a single endogenous explanatory variable, a F-statistic below 10 is of concern. All our F-statistics are far greater than 10.

Table 4.4 – Relevance of the lagged stocks of immigrants as instruments

Dependent variable:	ln(1+ Immigrants 1999)			
	(1)	(2)	(3)	(4)
ln(1+Immigrants 1990)	0.566^a	0.503^a	0.488^a	0.505^a
	(0.007)	(0.01)	(0.012)	(0.013)
ln(1+Immigrants 1982)		0.218^a	0.242^a	0.24^a
		(0.01)	(0.012)	(0.013)
ln(1+Immigrants 1975)			0.045^a	0.061^a
			(0.011)	(0.013)
ln(1+Immigrants 1968)				-0.012
				(0.011)
Distance	-0.055	0.106^b	0.155^a	0.146^a
	(0.047)	(0.041)	(0.038)	(0.036)
Contiguity	0.854^a	0.665^a	0.573^a	0.534^a
	(0.112)	(0.094)	(0.08)	(0.075)
Obs.	8011	5471	4038	3558
Adj. R^2	0.934	0.949	0.961	0.965
$F(N_1, N_2)$	6069.6	3969.4	2886.1	2285.2
N_1	1	2	3	4
N_2	7805	5306	3881	3400
BJB R^2	0.44	0.6	0.69	0.73

Notes: Country and *département* fixed effects are not reported here. Robust standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

fulfill the orthogonality conditions. Regarding exports, the statistic is equal to $\chi^2(2) = 0.45$ with a p-value at 0.8, whereas for imports, the value is $\chi^2(2) = 1.25$, with a p-value at 0.53. In both cases, we thus fail to reject the null hypothesis. The fail of the rejection of the null is a further proof of the validity of instruments.

Results from instrumented regressions

In the columns (1) and (5) of table 4.5, we estimate the log-linear specification for all the observations for which the stock of immigrants in 1990 is non-missing. This slightly reduces the sample. The pro-trade effect of immigrants is broadly the same for exports and imports, with an elasticity at 0.112. Doubling the stock of immigrants yields a trade increase of 8%. This is the new benchmark against which we assess the endogeneity bias.

In columns (2) and (6), we report the estimates drawn from the just-identified model. Instrumentation confirms the significant and positive impact of immigration on exports and imports. Even though the elasticities are slightly reduced, which means that benchmark estimates were plagued by a small upward endogeneity bias, the orders of magnitude remain fairly stable, around 0.095. To the best of our knowledge, no such a formal robustness check had been proposed in the literature.

Columns (3) and (7) provide OLS estimates for the log-linear specification, based on the country-pairs for which all past stocks of immigrants are non-missing. This reduces drastically the number of observations. However, instrumented regressions reported in columns (4) and (8) provide estimates that are not significantly different from OLS results. This confirms that, even on this small sub-sample, the positive impact of immigration on trade is not driven by a reverse causality or an omitted variable bias.

Table 4.5 – Instrumented regressions at the *département*-level

	Export				*Imports*			
	Just-identified		*Over-identified*		*Just-identified*		*Over-identified*	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Distance	-0.704^a	-0.711^a	-0.62^a	-0.62^a	-1.533^a	-1.541^a	-1.318^a	-1.312^a
	(0.083)	(0.083)	(0.074)	(0.072)	(0.128)	(0.127)	(0.117)	(0.115)
Contiguity	0.322^b	0.357^b	0.274^c	0.281^b	0.167	0.205	0.18	0.081
	(0.161)	(0.164)	(0.142)	(0.141)	(0.196)	(0.2)	(0.192)	(0.191)
Immigrants	0.115^a	0.094^a	0.162^a	0.159^a	0.12^a	0.099^b	0.186^a	0.239^a
	(0.018)	(0.026)	(0.021)	(0.025)	(0.029)	(0.041)	(0.035)	(0.042)
Obs.	7833	7833	4022	4022	7097	7097	3880	3880
Adj. R^2	0.854	0.854	0.882	0.882	0.809	0.809	0.843	0.843

Notes: Country and *département* fixed effects are not reported here. Robust standard errors in brackets with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

To sum up, immigrants do have a positive and significant impact on both exports and imports. A doubling of the stock of immigrants increases the value of exports by 7 to 12%, depending on the sample and the estimation procedure. The impact on imports, between 7 and 18%, is slightly more variable but of the same order of magnitude. We further find that these results are robust to specification and selection biases and that endogeneity introduces only a slight upward bias in OLS estimates.

4.4 Product complexity, quality of institutions and immigration

In this last section, we study the pro-trade effect of immigrtion along two intertwined dimensions: the degree of complexity (or differentiation) of traded products, and the quality of institutions in partner countries.

The complexity of traded goods

Rauch (1999) is the first to argue that trade impediments would depend on the degree of differentiability of traded products. He distinguishes differentiated goods from those sold on an organized market or possessing a reference price. In a gravity-type model of international trade, he provides convincing evidence that proximity, common language and colonial ties matter more for the former than for the latter. Using the same classification, Rauch and Trindade (2002) even argue that the trade-creating impact of immigration, the Chinese diaspora in their study, is much more salient for differentiated than for homogeneous goods. Hence, transnational networks would bridge the information gap between international sellers and buyers in a more salient way for trade in differentiated goods.

We investigate a similar conjecture for French *départements* and their international trade partners. We first match the NST/R industrial classification with the 4-digit SITC classification of Rauch.[12] We consider two types of goods only: simple and complex goods. Simple goods are either those exchanged on an organized market or those possessing a reference price. Complex goods are all the other ones, classified by Rauch as differentiated goods.[13] We estimate now:

$$\ln y_{kij} = f_{ki} + f_{kj} - \beta \ln \text{dist}_{ij} + \gamma \text{contig}_{ij} + \alpha_k \ln \left(1 + \text{mig}_{ij}\right) + \varepsilon_{kij}, \tag{4.5}$$

where k indices the type of goods, with $k \in (simple, complex)$. Exports and imports, as well as country and *département* dummies, are now commodity-specific. Whereas we assume that the distance and contiguity effects do not vary across goods,[14] the elasticity of trade with respect to the stock of immigrants is also commodity-specific. Contrary to Rauch and Trindade (2002), we run two separate regressions for exports and imports.

Table 4.6 – Product type and immigration

	Exports				*Imports*			
	In log		In levels		In log		In levels	
	OLS		2NB≥ 0		OLS		2NB≥ 0	
	Simple	*Complex*	*Simple*	*Complex*	*Simple*	*Complex*	*Simple*	*Complex*
Distance	-0.775^a (0.072)		-0.951^a (0.086)		-1.492^a (0.099)		-1.603^a (0.124)	
Contiguity	0.371^a (0.143)		0.19 (0.134)		0.425^a (0.155)		0.082 (0.181)	
Immigrants	0.141^a (0.025)	0.074^a (0.018)	0.123^a (0.025)	0.095^a (0.022)	0.029 (0.035)	0.075^a (0.027)	0.05 (0.044)	0.113^a (0.043)
Obs.	17711		18800		15396		18988	
Adj. R^2	0.809				0.766			

Notes: Country and *département* fixed effects are not reported here. Robust standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

Table 4.6 reports the OLS estimates for specification (4.5) in log (columns OLS) and the 2-step negative binomial QML estimates for specification (4.5) in levels (column 2NB≥ 0). A first striking feature is that the trade-creating effect of immigration is now different for exports and imports. Recall that, when the type of goods was not taken into account, the pro-trade effect of immigrants was of the same order of magnitude for exports and imports. By way of contrast here, immigration boosts the imports of complex commodities

[12]See appendix 4.7 for further details.

[13]Berkowitz et al. (2006) follow the same dichotomy. Results are not drastically changed if we consider three categories separately.

[14]Allowing these elasticities to be commodity-specific does not change the estimates of the impact of immigrants. However, it reduces the precision of the distance and contiguity estimates but, as noted above, this remains difficult to interpret.

(with an elasticity at 0.113), whereas it has no significant impact on the imports of simple products.[15] This is consistent with the idea that social networks, by providing market information and supplying matching or referral services, would matter more for the imports of complex products. Regarding exports, migrants have a significant impact on both simple and complex goods. The effect would be even slightly stronger for simple goods, even if the difference is not significant.

Such average elasticities could hide another source of heterogeneity, depending on the partner country characteristics, as recently suggested by Bandyopadhyay et al. (2008). In the following, we disentangle further the pro-trade impact of immigration according to the rule of law in partner countries, on aggregate flows first and then, by type of goods.

The quality of the trading partner's institutions

Some recent papers study the impact of institution quality on the volume of bilateral trade. In a matching model of international trade, Turrini and van Ypersele (2006) provide new evidence on the deterrent impact of legal asymmetries on bilateral trade between OECD countries, as well as between French regions. Besides, Anderson and Marcouiller (2002) establish that good institutions would reduce predation at the border. They find that a 10% rise in a country index of transparency and impartiality yields a 5% increase in its import volumes, other things equal.[16]

Berkowitz et al. (2006) add that the quality of the exporter's institutions matters even more. They argue that, if some common contracts (as letters of credit, counter-trade agreements and pre-payment) exist to offset the exporter's risk of not getting paid, such devices are scarcer to offset the importer's risk of late delivery and product defects. Therefore, formal institutions, such as courts and arbitration tribunals for seeking compensation, are of primary interest for importers. Most of the time, the courts or arbitration tribunals in the export country are indeed the last fallback for resolving disputes, the reason why the quality of institutions is more important in the export country.

Rauch (2001) puts forward the idea that transnational networks could be a substitute for weak institutions or weak mechanisms of arbitration. But, as far as we know, this effect has only been empirically studied by Dunlevy (2006), who restricts the focus to U.S state exports. We further investigate the conjecture of transnational network as a substitute for weak institutions on both the international exports and imports of French *départements*. According to Anderson and Marcouiller (2002), the impact of immigration should be greater for exports, as immigrants mitigate any predation behavior at the border of the importing country. According to Berkowitz et al. (2006), this should be the reverse as immigrants substitute for weak arbitration tribunals in the exporting country.

Crossing the effects of migrants and institutions may allow us to identity which one of the two previous views is the most salient. We use the *rule of law* index (hereafter RL) provided by Kaufmann, Kraay, and Mastruzzi (2007) as a measure of the quality of

[15] In the remaining, we comment the results associated with estimations in levels only, differences with estimates in logs being most of the time insignificant.

[16] See also de Groot, Linders, Rietveld, and Subramanian (2004) and Ranjan and Lee (2007).

institutions. This index measures "the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement, the police and the courts, as well as the likelihood of crime and violence". This variable is thus very close to the reality we want to describe.[17]

We proceed with the following estimation:

$$\ln y_{ij} = f_i + f_j - \beta \ln \text{dist}_{ij} + \gamma \text{contig}_{ij} + \alpha \ln \left(1 + \text{mig}_{ij}\right) + \rho RL_j * \ln \left(1 + \text{mig}_{ij}\right) + \varepsilon_{ij}, \tag{4.6}$$

where the (log of the) stock of immigrants is crossed with the RL index in country j (RL_j). In line with Rauch (2001), we conjecture that immigrants from partner countries with weak institutions have a larger impact on trade flows, in which case we expect a negative sign for ρ.

One could argue that the quality of institutions is endogenous to trade openness, and thus to the volume of trade. If this assertion is certainly right in general, we can forcefully argue that France remains a marginal trading partner for a large majority of countries in the sample. Hence, bilateral flows with France do not determine the quality of its trading partners' institutions. Moreover, the largest trading partners of France are high-income countries, where the quality of institutions is already high.

Table 4.7 – Immigration and the quality of the partner's institutions

	Exports		*Imports*	
	In log	In levels	In log	In levels
	OLS	2NB ≥ 0	OLS	2NB ≥ 0
Distance	-0.839^a	-1.014^a	-1.510^a	-1.678^a
	(0.086)	(0.108)	(0.127)	(0.16)
Contiguity	0.449^a	0.265	0.451^b	0.18
	(0.172)	(0.176)	(0.206)	(0.235)
Immigrants	0.085^a	0.096^a	0.047^c	0.078^c
	(0.018)	(0.02)	(0.027)	(0.04)
RL*Immigrants	-0.067^a	-0.053^a	-0.042^a	-0.058^a
	(0.009)	(0.013)	(0.014)	(0.02)
Obs.	9033	9400	8110	9494
Adj. R^2	0.845		0.8	

Notes: Country and *département* fixed effects are not reported here. Robust standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

Table 4.7 reports the estimates of specification (4.6). Note first that the direct trade-impact of institution quality is captured by the country-specific dummy and thus, it cannot be separately identified. Due to the normalization of the rule-of-law index to a zero mean,

[17]Kaufmann et al. (2007) provide six different measures of the quality of institutions. Due to the strong correlation between these measures, we restrict the focus to the rule-of-law index. However, results are unchanged when another index is chosen. The index is decreasing in the quality of institutions and stands between -2.5 and 2.5. We proceed to a simple normalization so that our sample mean would be zero and standard deviation would be one.

the average impact of immigrants is taken into account via the *Immigrants* variable. It is almost the same as in section 4.3. The interacted term *RL*Immigrants* accounts for an heterogeneity in the immigrant effects that depends on institution quality in partner countries. Our results support the conclusion of Dunlevy (2006). The coefficient is negative for exports: immigrants matter more when the quality of institutions is weak in the home country. We compute that the elasticity of exports to immigration ranges between 0.16, for the country with the lowest rule of law (Congo) to an insignificant 0.01 for the country with the highest value (Netherlands).

In addition to Dunlevy (2006), we also provide the related estimates for imports. The impact of immigration also presents a high heterogeneity. The elasticity ranges from 0.15 for the first decile of institution quality to a zero effect for the last decile. Finally, the above-mentioned mechanisms by which weak institutions could impact on trade flows are not exclusive. However, immigrants mitigate the trade-reducing impact of weak institutions in both directions.

Complex products, quality of institutions and immigration

According to our previous discussion, the pro-trade effect of immigrants depends on both the type of goods and the quality of institutions. Hence, it makes sense to study the triple interaction. In the following, we evaluate the cross effect of institutions and immigrants for simple and complex goods separately. Results are reported in table 4.8.

Table 4.8 – Product type, quality of institutions and immigration

	Exports				*Imports*			
	In log		In levels		In log		In levels	
	OLS		2NB≥ 0		OLS		2NB≥ 0	
	Simple	*Complex*	*Simple*	*Complex*	*Simple*	*Complex*	*Simple*	*Complex*
Distance	-0.856^a		-1.008^a		-1.527^a		-1.654^a	
	(0.072)		(0.089)		(0.098)		(0.126)	
Contiguity	0.601^a		0.389^a		0.554^a		0.299	
	(0.151)		(0.143)		(0.16)		(0.183)	
Immigration	0.118^a	0.058^a	0.107^a	0.084^a	0.023	0.07^a	0.038	0.106^b
	(0.025)	(0.018)	(0.026)	(0.022)	(0.035)	(0.027)	(0.044)	(0.042)
RL*Immigration	-0.111^a	-0.065^a	-0.075^a	-0.05^a	-0.08^a	-0.023^c	-0.116^a	-0.024
	(0.013)	(0.01)	(0.015)	(0.012)	(0.019)	(0.013)	(0.02)	(0.021)
Obs.	17711		18800		15396		18988	
Adj. R^2	0.806				0.766			

Notes: Country and *département* fixed effects are not reported here. Robust standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

For exports, immigrants enhance trade for both types of goods, even more when the quality of institutions is low, which matches aforementioned intuitions. However, the direct effect is slightly stronger and more heterogenous across rules-of-law for simple goods.

Regarding the imports of complex goods, the role of immigrants does not depend on the quality of institutions. Since for complex goods immigrants are a real conduit for information, they matter regardless of institution quality. For simple goods conversely, immigrants do not matter on average, because trading such goods does not require further information enhancement: hence, the direct effect is not significant. This result holds unless the quality of institutions is low. In that case, immigrants, who substitute for institutions, play an important role, as shown by the negative significant effect of the interacted variable.

4.5 Conclusion

The positive impact of immigration on trade is a well-established result. We add to the literature by assessing the cross-effect of immigration, goods complexity and institution quality. Even though numerous theoretical models underline this possible interaction, evidence remains very scarce.

When we do not disentangle the pro-trade effect of immigrants across goods and institutions, we find that the trade-creating impact of immigrants is slightly smaller than that found in the previous literature. This might be due to our careful estimation strategy, in which we consider variables in levels, country fixed-effects and instrumentation. However, these average effects hide a large heterogeneity across products and across trading partners.

The trade-enhancing impact of immigrants is more salient when they come from a country with weak institutions. Doubling the stock of immigrants from countries with the weakest institutions increases exports and imports by 10 to 12%. Conversely, the impact of immigrants is barely significant for countries with best institutions.

Furthermore, immigrants substitute for weak institutions for the exports of both simple and complex goods. Regarding the imports of complex commodities, i.e. those for which the information conveyed by immigrants is the most valuable, the pro-trade effect of immigrants overrides institution quality in the partner country. Conversely, even though immigrants do not enhance the imports of simple goods on average, they play an important role in interaction with the quality of institutions.

4.6 Appendix A to chapter 4: Data on trade and immigration

Trade flows

Trade flows come from the SITRAM dataset provided by the French Ministry of Transport. It reports the value of imports and exports of 94 French metropolitan *départements* with around 200 trading partners all around the world. French *départements* are administrative units of much smaller and more regular size than US States or Canadian Provinces. The mean area of French *départements* is 5,733 km^2, with a coefficient of variation at 0.34 (when Corsica and overseas French regions are excluded), whereas the related figures are 162,176 km^2 (with a standard deviation at 0.77) for US states (when Alaska and Washington DC are included), and 606,293 km^2 (with a standard deviation at 0.82) for Canadian provinces (when Nunavut, North-West and Yukon territories are excluded).

These flows are available for the years 1978 to 2002. However, the set of countries fluctuates over time. The instrumentation strategy requires that countries remain comparable across time. And the decade 1990-2000 has seen a large deal of modifications in the drawing of countries with, for instance, the disaggregation of the former Soviet Union and of Ex-Yugoslavia. Hence, we recover those entities as they were before the separation:

- Four former single countries have been divided during the 1990's. In order to match the data set in 1999 with our explanatory variables, we thus aggregate Armenia, Azerbaidjan, Belarus, Estonia, Georgy, Kazakhstan, Kirghistan, Lettonia, Lituania, Moldova, Ouzbekistan, Russia, Tadjikistan, Turkmenistan and Ukrainia in a single *former Soviet Union*. Czech Republic and Slovakia are aggregated in *former Czecholovakia*, Bosnia, Croatia, Serbia, Montenegro, Slovenia and Macedonia in *former Socialist Republic of Yugoslavia*, Erythrea and Ethiopia in *former Ethiopia*.

- We also aggregate three countries that have been reunified during the 1990's: Germany (former DDR and former GDR), Yemen (former South and North Yemen), and the Emirates.

We further consider as a single country: 1/Belgium and Luxembourg, 2/Italy, San Marin and Vatican, 3/Denmark and Feroe Islands, 4/Switzerland and Lichtenstein. After this manipulation, 161 countries remain in the data set, with at least one positive flow towards or from a French *département*.

As noted in the main text, the value of trade flows is generally exclusive of transit shipments. Petroleum products are however a noticeable exception. Hence, we leave them out of the sample. We also neglect postal, pipers and other too specific shipments.

The distributions of exports and imports across countries are right-skewed, with a set of few countries accounting for the largest amount of trade flows: nine countries only account for more than 70% of the value of exports and of imports (Germany, Belgium/Luxembourg, Spain, Italy, the Netherlands, United-Kingdom, United-States, Switzerland and Japan). It is also worth noting that half of the sample (80 countries) accounts for 98% (99%) of the value of exports (imports). Furthermore, import and export countries are very similar: the Spearman rank correlation between importers and exporters stands at 0.86.

Immigration

The 1999 French population census, from the French National Statistical Institute (INSEE), provides us with exhaustive information on the number of foreign-born residents by *département*. For each foreign-born resident, we know the country of birth, the nationality at birth, and the nationality at the time of the census. We are then able to distinguish between 1/French citizens born abroad, 2/foreign citizens born in France, 3/foreign citizens born abroad but having acquired the French nationality, and finally 4/foreign citizens born abroad with a foreign nationality at the time of the census.

As the place of birth is more important in the construction of a social network than the current nationality, we consider the narrower concept of *immigrant*. The French Statistical Institute disentangles a *foreigner*, i.e. a person whose current nationality is not French, from an *immigrant*, i.e. a person born abroad with a foreign nationality, regardless of his/her nationality at the time of the census. Hence, if an immigrant acquires the French nationality, he/she cannot be considered a foreigner anymore, but remains an immigrant. Note that for a few countries, it is necessary to sort apart French citizens born abroad from foreign-born French citizens. The Algerian case is very enlightening in this respect. Eighteen French *départements* count more than 10,000 French citizens born in Algeria, who are not immigrants (Algeria was a settlement colony of France until 1962). The settlement pattern of French citizens born in Algeria and Algerian-born citizens is not completely similar, with a correlation at 0.64 only.

The distribution of immigration across countries is also highly right-skewed. Eight countries account for more than 70% of immigrants to France (Algeria, Morocco, Portugal, Italy, Spain, Tunisia, Germany and Turkey). Most of these countries do not stand in the top-9 French trading partners. The geography of trade and immigration is thus quite different. The correlation between immigration and exports (imports) stands at 0.65 (0.56). This correlation is only 0.22 (0.20) when we restrict the sample to countries belonging to the upper-median part of the distribution.

To prevent the results from being driven by noisy observations and the skewness of our three variables of interest, we restrict the sample of exports, imports and immigration stocks to the upper-median distribution countries. This leads us to consider a sample of 100 countries for exports and a sample of 101 countries for imports.

Description of the instruments

The French population censuses of 1968, 1975, 1982 and 1990 provide us with a further reliable information on the number of immigrants by *département* and by country of origin, used as instruments to tackle the endogeneity issue. It is worth noting that, for earlier censuses (1968 and 1975), information is not exhaustive as it is extracted from a representative sample (1/4 of the whole French population). Moreover, for these years, we only know the nationality of the residents (and not the country of birth) for a limited number of countries. Hence, the number of observations reduces drastically when we use these variables as instruments. The 1982 and 1990 censuses provide the nationality of the respondent, as well as his/her country of birth. We are then able to recover an instrument variable closer to the endogenous explanatory variable.

Summary statistics

Table 4.9 depicts further summary statistics on the distributions of exports, imports, distance and immigration over the *département*-country pairs. In the panel of exports, there are 9033 pairs (among 9400 possibilities) of strictly positive flows, against 8110 (among 9494 possibilities) for imports, with a slightly greater pair-average value (31,980 thousands of euros against 30,443 for exports). The frequency of null flows is then quite limited here, in comparison to Helpman, Melitz, and Rubinstein (2008) for instance (half of the sample).

Table 4.9 – Summary statistics

	Mean	Std. Dev.	Min	Q25	Q50	Q75	Max
Strictly positive exports (9033/9400)							
Exports	30,443.2	134,961.7	0.2	311.4	2,122.5	12,621.7	3,500,597.5
Distance	5,321.9	3,758.0	110.6	1,956.8	4,608.3	8,358.1	19,839.1
Immigrants	470.6	2,224.0	0.0	7.0	29.0	140.0	56,540.0
All exports (9400)							
Exports	29,254.6	132,431.9	0.0	234.1	1,848.3	11,694.1	3,500,597.5
Distance	5,338.8	3,712.9	110.6	2,021.2	4,638.2	8,325.4	19,839.1
Immigrants	452.8	2,181.9	0.0	6.0	27.0	131.0	56,540.0
Strictly positive imports (8110/9494)							
Imports	31,079.7	151,225.4	0.1	54.9	890.0	9,076.8	4,451,061.5
Distance	5,626.0	3,933.6	110.6	1,912.3	4,983.7	8,908.9	19,839.1
Immigrants	519.2	2,341.7	0.0	7.0	34.0	170.0	56,540.0
All imports (9494)							
Imports	26,549.0	140,197.3	0.0	7.2	392.2	6,335.9	4,451,061.5
Distance	5,577.7	3,704.2	110.6	2,238.5	4,954.2	8,615.1	19,839.1
Immigrants	448.1	2,171.6	0.0	5.0	26.0	128.0	56,540.0

Notes: Exports and imports are in thousands of euros, immigrants in number of foreign-born French residents. Distance is the average number of kilometers between capital cities, weighted by their population size.

4.7 Appendix B to chapter 4: Matching the NST/R and Rauch's classifications

The NST/R classification consists in a 3-tier nomenclature: 10 *chapters*, 52 *groups*, and 176 *positions*. We match each of these *positions* with the nomenclature built by Rauch (1999), who classifies the 1089 goods of the 4-digit SITC (rev. 2) system into three broad categories: the goods sold on an organized market, the reference price goods or neither of the two. Rauch (1999) provides a conservative and a liberal classification. In the main text, we use the conservative one, but we check that the results are not sensitive to the alternative classification. We cannot define a one-to-one mapping between the categories of Rauch, and the NSTR classification. Therefore, we measure how each *position* distributes across these three broad categories.

To this aim, we use a correspondence between the 6-digit Harmonized Standard (HS6) and the NST/R classifications on one side, and between the HS6 and the classification of Rauch (1999) on the other side. The distribution of each *position* across the three Rauch's categories is computed as the ratio of the number of HS6 items belonging to each category over the number of HS6 items composing a given *position*.

To compute a correspondence table between the NST/R and HS6 classifications, we first use the correspondence table between the 8-digit Combined Nomenclature (CN8) and the NST/R classifications provided by the European Statistical Institute (EUROSTAT). We then use another correspondence table provided by EUROSTAT for the year 1988 to match each CN8 item with only one item of the HS6 classification.

In order to compute a correspondence between the HS6 and the classification of Rauch (1999), we use a correspondence table between the 4-digit SITC (rev. 2) and the 10-digit Harmonized Standard (HS10) classifications provided by Feenstra (1996).

Table 4.10 provides the distribution of each NST/R *chapter* across the three broad categories defined by Rauch. As expected, differentiated goods mainly appear in chapter 9 (Machinery, transport equipment, manufactured articles), and homogeneous goods in chapters 0 and 4.

Table 4.10 – Distribution of the 9 NST/R chapters across Rauch's categories (in %)

Chapters	Label	n	r	w
0	Agricultural products and live animals	19.69	25.87	54.44
1	Foodstuffs	19.26	67.6	13.13
2	Solid mineral fuels	13.77	86.23	0
4	Ores and metal waste	0	60.54	39.46
5	Metal products	29.91	63.56	6.53
6	Crude and manufactured minerals	66.6	33.4	0
7	Fertilizers	3.82	96.18	0
8	Chemicals	59.42	40	0.58
9	Machinery, transport equipment and manufactured articles	96.5	3.17	0.34

Notes: n = Differentiated Goods, r = Reference Price Goods, w = Goods sold on an organized market. Chapter 4 (petroleum products) is left out of the analysis.

4.8 Complementary tables

The first column of table 4.11 reports OLS estimates equivalent to those presented in table 4.2. The second column, $OLS(y + 0.1)$ gives the related estimates for the log-linearized model, where the dependent variable has been replaced by the logarithm of 0.1 plus the flow (in thousands of euros). This methodology has been used by Dunlevy (2006), Bénassy-Quéré, Coupet, and Mayer (2007) among others. The third column ($ET-Tobit$) gives the gravity estimates building on a modified Tobit estimator, as suggested by Eaton and Tamura (1994). This method has been used by Herander and Saavedra (2005).

The three following columns report QML estimates. The first column ($2NB$) depicts the results of a 2-step Negative Binomial procedure similar to that of table 4.2. The second column ($GPML$) presents another QML estimator, where we assume that the error term follows a Gamma distribution. The third column ($PPML$) depicts the Poisson QML estimates used by Santos Silva and Tenreyro (2006).

Table 4.11 – Results from different specifications

	In *Log*			In *Levels*		
	OLS	OLS($y+0.1$)	ET-TOBIT	2NB	GPML	PPML
Exports (>0)	0.102^a	0.101^a	0.082^a	0.092^a	0.091^a	0.24^a
	0.018	0.018	0.014	0.019	0.019	0.035
Exports (≥ 0)	–	0.135^a	0.077^a	0.109^a	0.113^a	0.241^a
		0.021	0.013	0.021	0.021	0.035
Imports(>0)	0.054^b	0.055^b	0.068^a	0.094^a	0.095^a	0.208^a
	0.027	0.026	0.024	0.035	0.035	0.035
Imports(≥ 0)	–	0.032	0.057^a	0.089^b	0.120^a	0.208^a
		0.027	0.021	0.041	0.047	0.035

Notes: Robust standard errors in brackets, with ^{a}, ^{b} and ^{c} denoting significance at the 1%, 5% and 10% levels respectively.

CHAPTER 5

Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations?[1]

5.1 Introduction

Most empirical work in economic geography relies on scattered geo-coded data that are aggregated into discrete spatial units, such as cities or regions. However, the aggregation of spatial dots into boxes of different size and shape is not benign regarding statistical inference. The sensitivity of statistical results to the choice of a particular zoning system is known as the Modifiable Areal Unit Problem (hereafter MAUP). Surprisingly, economists paid little attention to this problem up until recently.[2] Our main objective here is to assess whether differences in results across empirical studies are really sparked by economic phenomena in the process under scrutiny, or rather just by different zoning systems. We first investigate whether changes in either the *size* (equivalently the number) of spatial units, or their *shape* (equivalently the drawing of their boundaries) alter any of the estimates that are usually computed in the economic geography literature. Second, we address the important question of whether distortions due to the MAUP are large compared to those resulting from specification changes.

Disentangling these two effects is essential for policy. For instance, much work has tried to check empirically whether agglomeration enhances economic performance at the scale of countries, European regions, U.S. states or even smaller spatial units such as U.S. counties or French employment areas. The magnitude of the estimates differs between papers, but we do not know whether this reflects zoning systems or real differences in the extent of knowledge spillovers, intermediate input linkages, and labor-pooling effects on firm productivity. The resulting economic policy prescriptions regarding cluster-formation strategies will be affected accordingly. In the same vein, a large body of literature has evaluated the degree of spatial concentration, but does not check whether the conclusion that some industries are more concentrated than others results from the chosen zoning system or from more fundamental differences in the size of agglomeration and dispersion forces across industries at different spatial scales.

[1]This paper is a joint work with Pierre-Philippe Combes (Univ. of Aix-Marseille & GREQAM) and Miren Lafourcade (Univ. of Paris 11- ADIS & PSE), forthcoming as Briant, Combes, and Lafourcade (2010).

[2]Two noticeable exceptions are Holmes and Lee (forthcoming) and Menon (2008).

This paper is based on three standard empirical questions in economic geography, although many others could have been considered.[3] We start by evaluating the degree of spatial concentration under three types of French zoning systems (administrative, grid and partly random spatial units) and by comparing the differences between concentration measures (Gini vs. Ellison and Glaeser) with those between zoning systems. We then turn to regression analysis as not only is the measure of any spatial phenomenon likely to be sensitive to the MAUP, but also its correlation with other variables. We estimate the impact of employment density on labor productivity and compare the magnitude of agglomeration economies across zoning systems and econometric specifications. Finally, we run gravity regressions. We study how changes in the size and shape of spatial units affect the elasticities of trade flows within France with respect to both distance- and information-related trade costs.

All of these empirical exercises suggest that, when spatial units remain small, changing their size only slightly alters economic geography estimates, and changing their shape matters even less. Both distortions are secondary compared to specification issues. More caution should be warranted with zoning systems involving large units, however. The MAUP is obviously less pervasive when data variability is preserved from one scale to another. When moving from dots to boxes, specific attention should be devoted to the following key points: 1 - the size of boxes in comparison with the original dots, 2 - the way data are aggregated, i.e. averaging or summation, 3 - the degree of spatial autocorrelation in the data. The MAUP is less jeopardizing when data are spatially-autocorrelated and averaged, as is the case in wage regressions. By way of contrast, the MAUP is more challenging when variables in a regression are not computed under the same aggregation process. In gravity regressions for instance, moving from one scale to another requires a summation of trade flows on the left-hand side, whereas distance is averaged on the right-hand side.

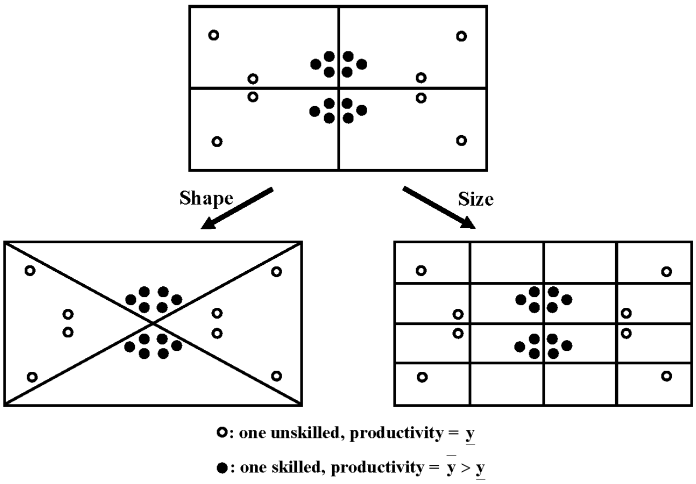
The remainder of the paper is organized as follows. Section 5.2 provides a simple illustration of the possible size- and shape-dependency of spatial statistical inference, along with a data simulation exercise. Section 5.3 lists the zoning systems for which our estimations are carried out. As a first sensitivity test, section 5.4 is dedicated to the study of French spatial concentration patterns. Sections 5.5 and 5.6 investigate the extent to which changing econometric specifications and zoning systems affect the size and significance of wage and trade determinants respectively. Section 5.7 concludes and suggests further lines of research.

5.2 The Modifiable Areal Unit Problem : A Quick Tour

The Modifiable Areal Unit Problem is a longstanding issue for geographers. In their seminal contribution, Gehlke and Biehl (1934) were the first to emphasize that simple statistics such as correlation coefficients could vary tremendously across zoning systems.

[3] For comparison purposes, we use the same specifications as those typically found in the literature (see Combes et al., 2008a), even though we do not necessarily think that they are the most apt.

Figure 5.1 – The size and shape issues



They note that, in the United States, the correlation between male juvenile delinquency and the median equivalent monthly housing rent increases monotonically with the size of spatial units. Openshaw and Taylor (1979) pursued this line of investigation and, drawing on correlations between the percentage of Republican voters and the percentage of the population over 60, standardize what they called the "Modifiable Areal Unit Problem".[4]

5.2.1 A simple illustration of the MAUP

Spatial statistics may vary along two dimensions: firstly, the level of aggregation, or the *size* of spatial units, and secondly, at a given spatial resolution, the drawing of their boundaries, or their *shape*. Figure 5.1 illustrates these two related issues via the employment density-labor productivity relationship. Black points display the location of skilled workers, whose individual productivity is denoted $\overline{y}$, while empty dots stand for unskilled workers, with productivity $\underline{y} < \overline{y}$. In the top figure, space is divided into four rectangles, each consisting of three skilled and two unskilled workers. The spatial distribution of workers across units is uniform and average productivity is the same across units. To illustrate the shape effect, consider the bottom-left figure. Spatial concentration emerges here, with two clusters of six high-skilled workers and two clusters of four low-skilled workers. Average productivity is higher in the former due to the spatial sorting of labor skills. Hence, agglomeration economies, defined here as the positive correlation between productivity and employment density, are zero in the first zoning system but positive in the second. We now turn to the size effect. In the bottom-right figure, we consider smaller rectangles with the same proportions as in the top figure. Spatial concentration is also found here, but the relationship between productivity and density is less marked than in the bottom-left case. Indeed, the difference in productivity between low- and high-productivity

[4]See Fotheringham and Wong (1991) for an extended review of the earliest MAUP contributions.

regions remains the same (except for empty boxes), whereas the density gap is higher in the bottom-right case. Hence, the extent and scope of agglomeration economies change with the size and shape of units, even though the underlying spatial information - the location and productivity of workers - remains the same.

The question we pursue in this paper is hence twofold. How much does moving from a particular zoning system to another alter the perception of an economic phenomenon? And how does this alteration vary accordingly to whether information is summed or averaged under this aggregation process? Section 5.2.2 provides a first clue to these questions, drawn from a simple simulation exercise.

5.2.2 Mean and variance distortions: a first illustration with simulated data

A number of authors have provided detailed analyses of the MAUP based on simulated data. According to Arbia (1989), both size and shape distortions are minimized (although never eliminated) under two restrictive conditions that are rarely met in practice: the exact equivalence of sub-areas (in terms of size, shape and neighboring structure) and the absence of spatial autocorrelation. In a subsequent work, Amrhein (1995) carries out a simulation exercise where he draws $10,000$ values from a randomly-generated variable and allots *randomly* each of these values to a Cartesian address within a unit square. In doing so, the value at one address is independent of the values at contiguous addresses and there is no spatial autocorrelation.[5] The author then divides the unit-square into, respectively, 100, 49 and 9 equally-sized sub-squares. Finally, he aggregates the information by averaging the values assigned to each sub-square. In line with Arbia (1989), he concludes that, under the strong assumption of random allocation, means do not display any pronounced size and shape effects and the changes in variances are only driven by the fall in the number of units.[6] Based on Canadian Census data, Amrhein and Reynolds (1997) further show that the distortions of simple statistics, such as the mean and variance, do not only depend upon the spatial organization of raw data, as reflected for example in their spatial autocorrelation coefficient, but also on the aggregation process, namely on whether information is either averaged or summed.

To get insights from more realistic data configurations, let us extend this literature and compare the distortions arising from both a random and a sorted process of spatial assignment of simulated data. Consider a unit segment with $10,000$ equally-spaced addresses.[7] Each address is given the occurrence of a log-normally-distributed variable.[8] To study size distortions, we aggregate the addresses so as to form spatial units that constitute a parti-

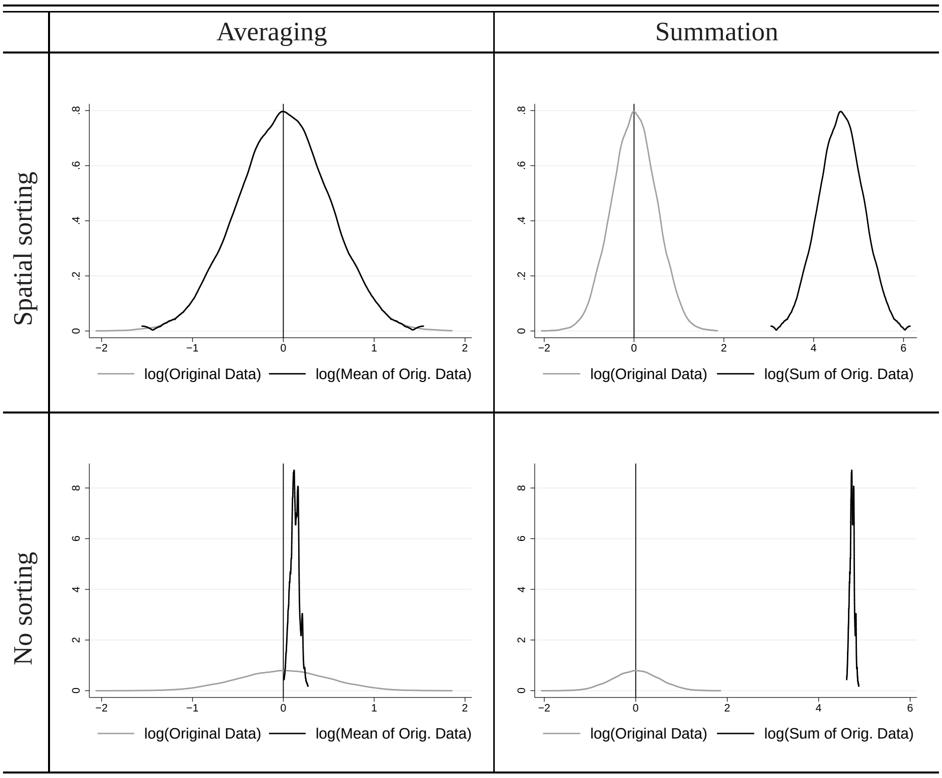
[5] More technically, Amrhein (1995) considers that the Cartesian coordinates of addresses are distributed either uniformly or normally, and that the generated variable follows either a normal or an uniform distribution.

[6] Under the same assumption of randomness, Holt, Steel, Tranmer, and Wrigley (1996) are able to justify theoretically the findings of Amrhein (1995). Note that Reynolds (1998) generates more realistic data configurations allowing for spatial autocorrelation.

[7] A two-dimension analysis of the MAUP would be more informative, but it is largely beyond the scope of the paper.

[8] The logarithm of the variable has a mean equal to 0 and a variance equal to 1.

Figure 5.2 – Aggregation with identically-shaped spatial units



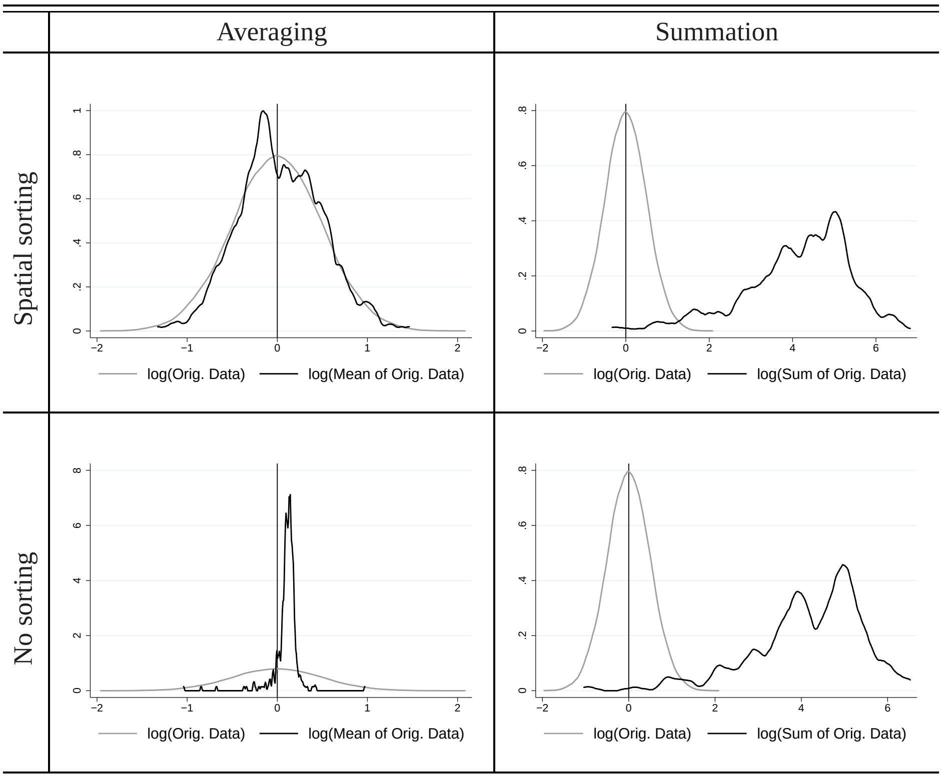
tion of the unit-segment. First, we choose equally-shaped spatial units. Then, we consider randomly-shaped spatial units, that do not include the same number of addresses. To see whether size distortions depend on how information is aggregated, we study four polar cases: data summation or averaging over the addresses of each spatial unit, with either perfect or no data sorting over address values. In the unsorted configuration, the value at a given address is independent of surrounding addresses, as is the case in Amrhein (1995). In the perfectly sorted configuration, the addresses are ranked by increasing order of their assigned values before aggregation. Figure 5.2 compares the log-distribution of the simulated data (tight line) with their log-distribution when spatial units are equally-shaped (thick line).[9] Three main conclusions emerge:

1. Mean and variance can be almost perfectly recovered after aggregation when data are spatially sorted, regardless of whether data are averaged or summed (top graphs). The support of the distribution is only slightly reduced after aggregation. In the case of summation, the distribution is shifted to the right by a constant that depends on the number of aggregated addresses.

2. More information is lost when data are not sorted. While the mean is more or less correctly inferred after aggregation (up to the above constant), the variance is greatly

[9] We define units such that they include 100 contiguous addresses.

Figure 5.3 – Aggregation with randomly-shaped spatial units



reduced (by a 10-fold factor with our parametrization).

3. In any case, the distribution form remains more or less the same, and keeps its single peakness.

Subsequently, with low *within-unit* heterogeneity (e.g. spatial sorting) and low *between-unit* heterogeneity (e.g. identically-shaped units), the first moments of the distribution are not too much distorted by aggregation and changes in the size of units. By way of contrast, with strong *within-unit* heterogeneity (e.g. unsorted data), aggregation yields a loss of information, even if units are shaped homogeneously.

Figure 5.3 shows that aggregation is likely to raise more concerns when spatial units are randomly-shaped:

1. When data are both sorted and averaged (top left graph), information can be partially recovered.

2. This is not the case anymore when data are unsorted (bottom left graph). As before, the variance is drastically reduced.

3. Summation is more problematic with randomly-shaped units, even if data are perfectly sorted (top right graph): it does not only shift the distribution to the right (so

as equally-shaped units), but it also enlarges the distribution support, thereby yielding an increasing dispersion of the variable.

To put it in a nutshell, when spatial units do not have the same shape, averaging is less sensitive to changes in size than summation, though part of the information is lost when data are not spatially-sorted. Conversely, if spatial units are randomly-shaped, summation is more distorted by a shift in their size. Distortions are even worse than data are unsorted.

5.2.3 Correlations distortions

Clear theoretical underpinnings are more difficult to come by for correlations, would they be univariate or multivariate. Fotheringham and Wong (1991), who consider a multivariate analysis of the determinants of mean household income for various zoning systems, come to an alarming conclusion: "The MAUP [...] is shown to produce highly unreliable results in the multivariate analysis of data drawn from areal units". They also find a sizable range for correlation and regression coefficients, which are positively (or negatively) significant for certain data configurations, but insignificant for others, suggesting that correlation inference is not robust to the aggregation process. Amrhein (1995) is the first to suggest separating aggregation effects from other types of discrepancies, such as model mis-specification in multivariate settings. In his simulation exercise, he shows that bivariate regression coefficients and Pearson correlations are sensitive to changes in the size and shape of spatial units, even if we know the data generation process and if we force the correlation between the two randomly-generated variables to be zero. However, he reaches a less alarming conclusion than Fotheringham and Wong (1991), and suggests that, for well-specified models, such as Amrhein and Flowerdew (1992), aggregation does not produce too many distortions, whereas for others, like Fotheringham and Wong (1991), the estimates are contaminated by size and shape.

Let us come back to our simulation exercise and turn to the analysis of regression coefficients. If aggregation distorts the explanatory and dependent variables in the same way, the size effect should be small. This is the case when, for instance, both the explanatory and dependent variables are spatially autocorrelated and averaged (top-left graph of figure 5.3). In sharp contrast, the size issue is more prevalent when the dependent and explanatory variables are not aggregated under the same process or do not exhibit the same degree of spatial autocorrelation.

As for shape distortion, it can be considered as a standard errors-in-variables issue. Let us consider the relationship $y^* = \beta_0 + \beta_1 x^* + \mu$, where y^* and x^* are two random variables, β_0 and β_1 two parameters, and μ an error term uncorrelated with x^*, and assume that the relationship is valid for a particular zoning system. Then, change the shape of spatial units so as to have $y = y^* + \varepsilon$ and $x = x^* + e$ for the new spatial units. It is straightforward to show that, under this new zoning system, regressing $y = \widetilde{\beta}_0 + \widetilde{\beta}_1 x + \nu$ gives a biased estimator of β_1:

$$\hat{\hat{\beta}}_1 = \beta_1 + \underbrace{\frac{cov(x^*,\varepsilon) - \beta_1 cov(x^*,e) + cov(e,\varepsilon) - \beta_1 \mathbb{V}[e]}{\mathbb{V}[x^*] + \mathbb{V}[e] + 2cov(x^*,e)}}_{bias}. \tag{5.1}$$

Note that there is no reason why the second right-hand term of equation (5.1) should be zero, except for knife-edge spatial configurations. Conversely, if the aggregation process generates random errors only and hence, $cov(x^*,e) = 0$, $cov(x^*,\varepsilon) = 0$ and $cov(e,\varepsilon) = 0$, the bias tends towards zero when $\mathbb{V}[x^*]$ grows faster than $\mathbb{V}[e]$. The larger the changes in borders, the larger the errors ε and e and thereby, the shape effect. Importantly, under the weaker condition that x is exogenous in the OLS regression of y on x, the bias is also zero.[10] In this respect, correcting the endogeneity of x, for instance with instrumental variables techniques, should alleviate the MAUP issue. Alternatively, improving specification should also reduce shape distortions, by making the explanatory variables more exogenous.

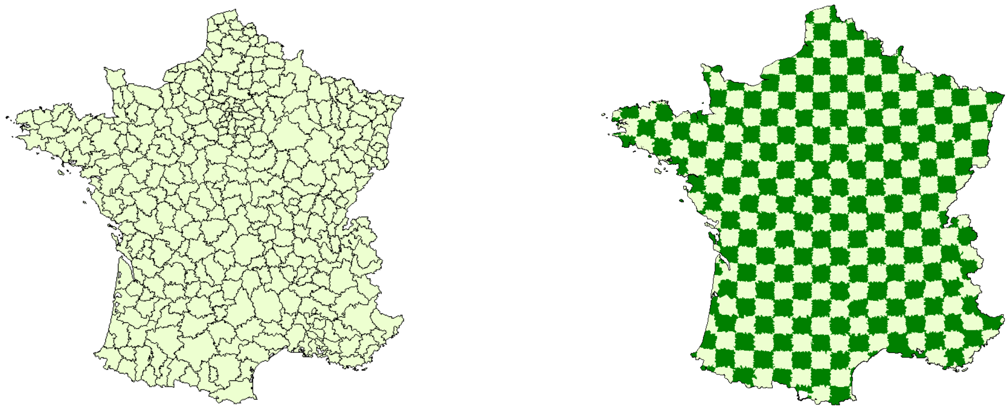
However, this exogeneity condition is not fulfilled if the value of x^* in one unit affects the outcome of the surrounding units (and therefore e, y^* and ε). The bias definitively increases with $cov(x^*,\varepsilon)$ and $cov(e,\varepsilon)$, i.e. with spatial correlation between x^* and y^*. By way of contrast, own spatial autocorrelation, reflected in $cov(x^*,\varepsilon)$, has a mixed effect on the magnitude of the bias. This is due to the spatial sorting effect highlighted in section 5.2.2, which mitigates the negative impact of non-random errors.

In what follows, we build on these intuitions to extend the MAUP literature in a number of ways. First of all, we systematically assess the magnitude of size and shape distortions relative to mis-specification biases. Secondly, we examine different aggregation processes to test the sensitivity of economic inference to the MAUP. In wage-density regressions, raw information is averaged over spatial units, while for gravity regressions it is either summed or averaged. In light of the above discussion, the former should be associated with less distortions than the latter and thereby, the distribution of wages and density variables should be barely unmodified by changing zoning systems. In contrast, the trade dependent variable might well experience an enlargement of its distribution support, whereas the dispersion of most of the trade explanatory variables should shrink. Therefore, MAUP distortions should be more salient in gravity regressions. Finally, we extend the work of Fotheringham and Wong (1991) by comparing the estimates from six different administrative and grid zoning systems to those from a hundred equivalent random systems.[11]

[10] We have $y^* = \beta_0 + \beta_1 x^* + \mu \Rightarrow y = \beta_0 + \beta_1 x + \mu + \varepsilon - \beta_1 e$. Variable x is exogenous if and only if $cov(x, \mu + \varepsilon - \beta_1 e = 0) \Leftrightarrow cov(x^*,\varepsilon) - \beta_1 cov(x^*,e) + cov(e,\varepsilon) - \beta_1 \mathbb{V}[e] = 0$.

[11] In this respect, our study echoes the work of Holmes and Lee (forthcoming), who investigate the prevalence of a Zipf's law for the U.S., based on an arbitrarily-drawn grid zoning system. It is also closely related to Menon (2008), who uses randomly-generated zoning systems equivalent to the commuting-defined Core Based Statistical Areas to study industrial agglomeration in the US.

Figure 5.4 – Small zoning systems



341 Employment Areas (EA) 341 Small squares (SS)

5.3 Zoning systems and data

The first zoning system we consider is that composed of 341 Mainland "Employment areas" (hereafter EA). These spatial units are underpinned by clear economic foundations, being defined by the French National Institute of Statistics and Economics (INSEE) so as to minimize daily cross-boundary commuting, or equivalently to maximize the coincidence between residential and working areas. This zoning system, currently composed of 341 areas, was designed to reduce the statistical artifact due to boundaries, which is why it is widely used in France. As can be seen on the left-hand side of figure 5.4, the average employment area is fairly small, covering 1570 km^2, which is equivalent to splitting the U.S. continental territory into over 4700 units.

Shape distortions can be identified from spatial units that are similar in size (or number) to employment areas. Conversely, size distortions can be highlighted with partitions of France involving units that are larger than the EAs. Hence, to disentangle the two faces of the MAUP, we appeal to three other sets of zoning systems.

5.3.1 Administrative zoning systems

The first set refers to French administrative units. Continental France is partitioned into 21 administrative "Régions" (RE), depicted on the left of figure 5.5, which are themselves split into 94 "Départements" (DE), shown on the left of figure 5.6. All such units are aggregates of municipalities, the finest spatial division for which data are available in France.[12]

It can nonetheless be argued that administrative boundaries do not capture the essence of economic phenomena that often spill over boundaries, which is one of the reasons why EAs were created. To circumvent this drawback, some authors, especially geographers,

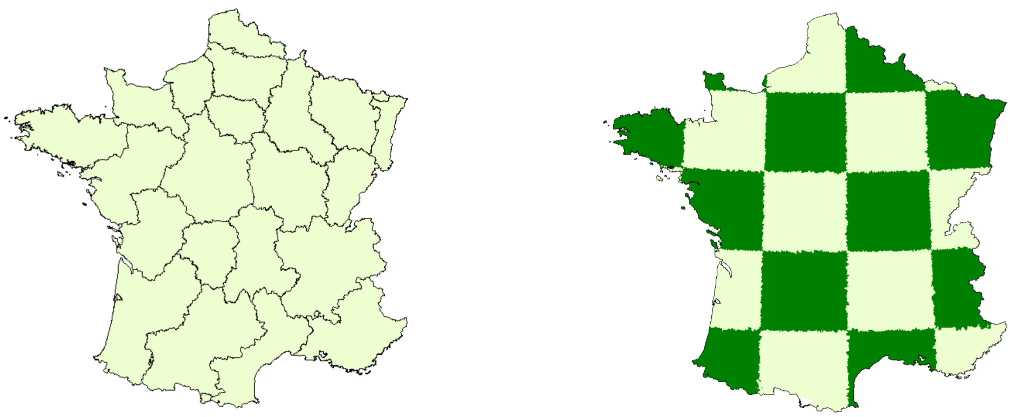
[12]The French metropolitan area is covered by 36,247 municipalities.

prefer to work with (often arbitrarily-drawn) checkerboard grids. The rationale is that, even if they do not necessarily better match the "true" boundaries of economic phenomena, grid zoning systems provide a greater degree of spatial homogeneity than do administrative zoning systems.[13]

5.3.2 Grid zoning systems

We therefore construct a second set of zoning systems purely based on grid units. We first enclose France into the smallest possible rectangle. We then divide this rectangle into lattices of squares (based on longitude and latitude). As France is more or less hexagonal, several squares jut out into the sea and we obviously left this out. We obtain the final grid by aggregating all municipalities which have their centroid into the same square. The resulting units are not perfect squares as their boundaries follow those of real municipalities. We choose the size of the squares to produce three different zoning systems analogous to administrative ones: 22 (non-empty) large squares (LS), 91 medium squares (MS) and 341 small squares (SS). It is worth noting that the largest zoning systems (LS and MS in figures 5.5 and 5.6) include several squares which are partially truncated due to French national boundaries. The finest grid such as SS (figure 5.4) circumvents this pitfall at the expense of geometry, since the units boundaries become increasingly ragged at the very fine scale. Therefore, overtly enlarging or tightening the units alters both their symmetry and regularity.

Figure 5.5 – Large zoning systems

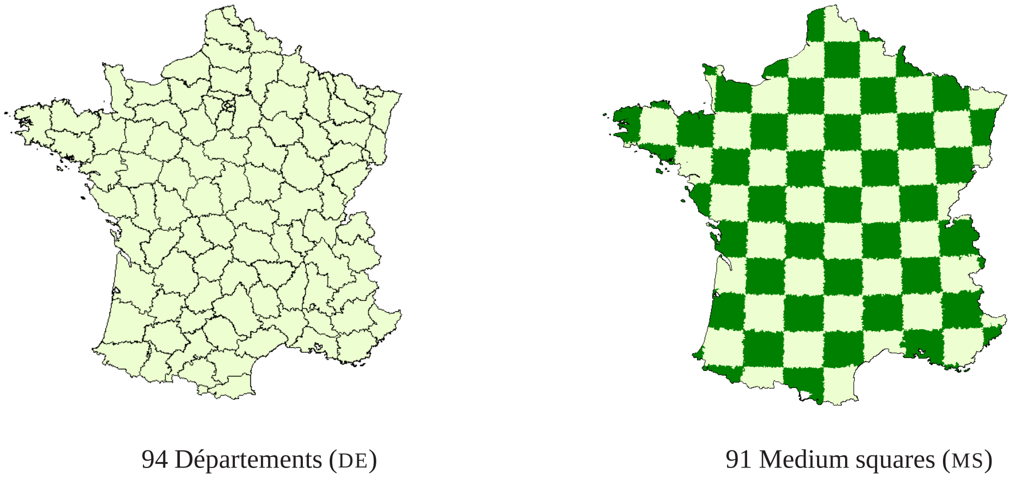


21 Régions (RE) 22 Large squares (LS)

A comparison of the results obtained under respectively RE, DE and EA or LS, MS and SS gives a flavor of any size distortions. We capture the impact of shape by comparing the results obtained across zoning systems involving units of similar size (RE to LS, DE to MS, and EA to SS). While these comparisons tell us whether MAUP distortions exist, they do

[13]Another argument is that grid zoning systems do not change over time, while administrative areas may do so. See ESPON (2006) for an overview of this issue.

Figure 5.6 – Medium zoning systems



not indicate whether the differences in the results are systematic and significant, however, which is why we propose a third set of zoning systems.

5.3.3 Partly random zoning systems

Our third set of zoning systems involves arbitrarily-drawn spatial units. We define a set of 100 different partitions of France, by randomly aggregating the 4662 French "Cantons",[14] into zoning systems that have a number of units strictly equivalent to those of administrative ones (341 units for EA, 94 for DE and 21 for RE): we call these REA, RDE and RRE respectively. These are constructed using the following algorithm. We randomly draw one canton, called the seed, within each administrative unit. We then aggregate each seed to a second canton randomly drawn from those contiguous to it. We continue with a third canton and so on, until all existing cantons have been drawn. We run the algorithm 100 times at each scale. Broadly speaking, this procedure produces, for each scale, a partition of France with jiggling borders.

5.3.4 Characteristics of zoning systems

Our empirical analysis builds on sectoral time-series data at the municipal level. The aggregation into the aforementioned larger zoning systems yields a three-dimension panel of employment, number of plants and wages for 18 years (within the 1976-1996 period) and 98 industries (at the two-digit level for both manufacturing and services). For 1996, we match this panel to a trade data set for manufactured goods.[15]

As can be seen in table 5.1, zoning systems differ sharply in their economic features. The spatial variation in land area is smaller for small grid units than for employment areas,

[14]We use this intermediate grouping of French municipalities to reduce the computational time without losing too much spatial variability in the randomization process.

[15]More details on the data are provided in 5.8.

Table 5.1 – Summary statistics

Zoning system		(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
Number of units		341	341	94	91	21	22
Land Area (km^2)	Av.	1569.8	1580.4	5733.3	5922.3	25663.4	24496.7
	Cv.	0.63	0.35	0.34	0.5	0.43	0.53
Employment (workers)	Av.	2012	2019	7300	7541	32678	31193
	Cv.	2.45	3.73	1.28	2.37	1.16	1.33
Employment density (workers/km^2)	Av.	4.6	1.5	12.3	1.7	1.8	1.3
	Cv.	8.7	3.1	6.3	1.7	1.8	0.8
Aggregate Market Potential	Av.	2910	2432	2956	2161	2137	1791
	Cv.	0.6	0.3	0.8	0.3	0.3	0.2
Municipality-level Market Potential	Av.	3300	2758	3585	2705	3097	2736
	Cv.	0.7	0.4	0.9	0.4	0.5	0.3
(Gross) Wage	Av.	1.3	1.2	1.3	1.3	1.3	1.3
	Cv.	0.2	0.2	0.1	0.1	0.1	0.1
Aggregate Distance (km)	Av.	393	419	384	445	371	448
	Cv.	0.49	0.48	0.5	0.49	0.52	0.52
Municipality-level Distance (km)	Av.	394	417	386	435	381	420
	Cv.	0.49	0.48	0.49	0.49	0.49	0.49
Trade Flow (tons $\times$ 1000)	Av.	91.05	102.22	382.83	496.8	5956.48	5839.38
	Cv.	6.6	6.5	5.6	6.2	3.1	3.2

Notes: (*i*) (EA): employment areas, (SS): small squares, (DE): Départements, (MS): medium squares, (RE): Régions, (LS): large squares. (*ii*) Averages over 18 years, except for trade flows (1996 value). (*iii*) Av. is the mean. Cv is the Coefficient of variation (standard deviation divided by mean). (*iv*) No unit for wage because detrended and centered around individual mean. No unit for market potential.

a property that does not hold for larger administrative units. This reflects two opposite effects. On the one hand, grid units are more regular, which reduces the variance. On the other hand, the share of truncated grid units increases with size, which increases the variance. The latter effect dominates for medium and large units. A clear drawback of the grid strategy is that, when units are not small enough, the gains of reducing the variance of land area cannot be attained due to the irregularity of national borders. Conversely, this also shows that the French authorities were fairly successful in designing quite homogeneous administrative units.

Regarding the other variables, an important distinction concerns the way in which information is aggregated. Some variables, such as employment and trade flows, are *summed*, whereas others, such as job density and wages, are *averaged*. The former increase with the size of the units, which is straightforward. By way of contrast, the overall picture vary less for averaged information. For instance, employment density differs only little across grid zoning systems, regardless of the size of their units, while it varies more for administrative units, which reflects that the design of administrative zoning systems was not based on this variable. Average wages are little affected by both administrative and grid zoning systems.

The suspicion that the MAUP could still bias the estimate of the impact of agglomeration economies motivates the exercise carried out in section 5.5.

However, there are two variables, distance and market potential, for which information is neither summed nor averaged. Consider first distance. It can be computed either as the great-circle distance between the centroids of spatial units ("Aggregate Distance" in table 5.1), or as the average distance between the municipalities of each unit ("Municipality-level Distance" in table 5.1). In the former case, there is no obvious link from one zoning system to the other, whereas in the latter, less information is lost through aggregation. The same argument holds for market potential. It can be the average of market potentials over municipalities or the aggregate market potential. Even if the two first moments of both couples of variables do not differ drastically, the MAUP could be more severe when variables are computed at the aggregate level. This source of distortions is investigated in sections 5.5 and 5.6.

5.4 Spatial concentration

Before turning to regression analysis, we carry out the most basic exercise in economic geography, which consists in measuring the extent of spatial concentration, an issue widely-covered in the literature. Apart from a small number of continuous approaches, such as Duranton and Overman (2005), work in this area is based on discrete zoning systems. While some work has focused on the comparison of spatial concentration across industries, such as Ellison and Glaeser (1997), only little has assessed the legitimacy of comparing results across zoning systems that differ in the size and shape of spatial units. In this section, we compare the variability in concentration due to the zoning system with that from different concentration indices.

5.4.1 Gini indices

We compute the spatial Gini index associated with every zoning system for 98 industries and 18 years (see 5.8). The moments of the index distribution are provided in table 5.2. Every moment of the distribution, in particular the mean, falls with aggregation level. The rationale is straightforward: smallers units have more areas with no registered employment for certain industries, which raises the Gini index mechanically for each industry.

We then rank industries by spatial concentration and compute Spearman rank correlations across zoning systems. The results are shown in table 5.3.

Rank correlations across zoning systems that are similar in size (EA and SS, DE and MS, and RE and LS) are very high, with values of at least 0.98 (see the sub-diagonal elements in table 5.3). The ranking of industries is therefore virtually unaffected by changes in the shape of units. Size has a slightly greater effect on concentration. For instance, the rank correlation between EA and RE is 0.95, which remains high. Making shape more homogeneous across scales leads to similar results, with the correlation between SS and LS zoning systems being 0.96.

Table 5.2 – Summary statistics for the Gini index

	Mean	St. Dev.	Min	P25	P50	P75	Max
(ZE)	0.587	0.224	0.134	0.410	0.597	0.767	0.994
(SS)	0.553	0.220	0.111	0.370	0.560	0.720	0.992
(DE)	0.481	0.217	0.098	0.299	0.465	0.637	0.980
(MS)	0.439	0.213	0.072	0.260	0.415	0.582	0.971
(RE)	0.338	0.187	0.051	0.184	0.321	0.443	0.947
(LS)	0.327	0.185	0.043	0.181	0.300	0.433	0.891

Note: Computed on 1764 observations (98 industries × 18 years).

Table 5.3 – Spearman rank correlations between Gini indices

Averages over 18 years

	(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
(EA)	1	0.99	0.99	0.99	0.95	0.95
(SS)		1	0.98	0.99	0.96	0.96
(DE)			1	0.99	0.97	0.97
(MS)				1	0.98	0.98
(RE)					1	0.98
(LS)						1

5.4.2 Ellison and Glaeser indices

It is well known that the spatial Gini index is contaminated by industry structure. Given total industry employment, industries with fewer plants will have higher Ginis, even with random plant location. Ellison and Glaeser (1997) develop a measure of concentration that is purged of this plant size effect. Table 5.4 describes moments of the EG index distribution.

Table 5.4 – Summary statistics for the Ellison-Glaeser index

	Mean	St. Dev.	Min	P25	P50	P75	Max
(ZE)	0.017	0.027	-0.015	0.004	0.009	0.019	0.396
(SS)	0.021	0.037	-0.065	0.004	0.012	0.027	0.365
(DE)	0.022	0.034	-0.014	0.005	0.012	0.025	0.407
(MS)	0.031	0.051	-0.067	0.004	0.014	0.039	0.364
(RE)	0.042	0.059	-0.062	0.006	0.023	0.051	0.434
(LS)	0.040	0.056	-0.116	0.005	0.018	0.052	0.326

Note: Computed on 1764 observations (98 industries × 18 years).

Contrary to the Gini coefficient, the EG index monotonically increases with the aggregation scale, which gives further support to well-known result already put forward by Ellison and Glaeser (1997), or Maurel and Sédillot (1999) and Devereux et al. (2004), for a slightly modified index. It can be taken as evidence that various industrial spillovers play at different scales. If we turn to the Spearman rank correlations, we have the results depicted in table 5.5.

Table 5.5 – Spearman correlations between EG indices

Averages over 18 years

	(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
(EA)	1	0.83	0.94	0.84	0.83	0.81
(SS)		1	0.79	0.87	0.85	0.84
(DE)			1	0.85	0.85	0.82
(MS)				1	0.93	0.90
(RE)					1	0.94
(LS)						1

The rank correlations are generally lower than those for the Gini indices. Hence, any distortions due to the MAUP are more pronounced when spatial concentration is measured via the EG index. In particular, size distortions are slightly aggravated, even though the rank correlations remain fairly high (0.83 for instance between EA and RE).

5.4.3 Comparison between the Gini and the EG

The success of the EG index over the Gini coefficient lies in its alleviation of concentration due to the location of big plants. In this respect, the EG index should be favored. The crucial question we address here is whether the zoning system affects the ranking of industries more than does the choice of the index itself. To answer, we turn to a between-index rank correlation analysis.

Table 5.6 shows that the between-index Spearman rank correlations are definitely smaller than their within counterparts. Even within each zoning system (the diagonal elements of table 5.6), the rank correlation is 0.81 at best (for RE), with the lowest correlation being 0.56 (for SS).

Table 5.6 – Spearman rank correlations between Gini and EG indices

Averages over 18 years

		Gini index					
		(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
EG index	(EA)	0.65	0.53	0.70	0.61	0.67	0.64
	(SS)	0.65	0.56	0.70	0.63	0.69	0.66
	(DE)	0.69	0.56	0.75	0.65	0.71	0.67
	(MS)	0.68	0.58	0.74	0.67	0.73	0.69
	(RE)	0.73	0.65	0.78	0.74	0.81	0.76
	(LS)	0.73	0.65	0.78	0.73	0.79	0.78

There is considerable evidence that index choice, which we can consider as a specification issue, produces greater distortions than the choice of zoning system, in terms of both size or shape. It should thus be of greater concern than the MAUP.

5.5 Agglomeration economies

While the MAUP only slightly distorts spatial concentration patterns, it might have a greater effect on the explanation of the spatial distribution of economic variables. We therefore now consider the incidence of the MAUP in the context of multivariate regression analysis. In this section, we focus on the estimation of agglomeration economies. Evaluating the magnitude of the benefits reaped from spatial proximity is important for policy, and much work, such as Ciccone and Hall (1996), has been devoted to the estimation of the productivity gains resulting from dense clusters of activities. The benefits from proximity to large markets and the local composition of labor skills are generally simultaneously estimated.[16]

We regress local wages, a frequently-used measure of local labor productivity,on local employment density. Let w_{at} denote the wage in area a at date t, computed as the *average* earnings of all workers located in a at date t (hereafter the "gross" wage), and Den_{at} employment density (per square-kilometer). The benchmark specification we run is the following:

$$\log w_{at} = \alpha \log Den_{at} + \gamma X_{at} + \varepsilon_{at}, \tag{5.2}$$

where X_{at} is a vector of control variables. We compare the estimated elasticity of wages to employment density across zoning systems. In this exercise, we consider the average wage and employment density per areal unit. In light of the simulations performed in section 5.2, we expect the MAUP to be mitigated in this setting. As for concentration indices, we then check whether the choice of zoning systems matters less for the magnitude of agglomeration economies than the biases from choice of controls in the wage equation, which is a specification issue.

5.5.1 A wage-density simple correlation

In order to have a benchmark, we first look at gross wage/density correlations. Given the panel structure of the data, we estimate equation 5.2 with no controls other than time dummies. Table 5.7 reports on the resulting elasticities.

The elasticity of wages with respect to employment density lies in the usual range of [0.04, 0.10] found for U.S. and European data (see Combes et al., 2008a). Even though some differences result from the move to a larger scale, the shape effect remains small.

Size differences do not really matter when moving from small to medium units, although larger differences occur as we move to the largest units. In both EA and DE, the value is about 0.07. However, the aggregation from DE to RE induces a 20%-increase in the coefficient estimate. As for the grid zoning system, the estimated elasticity is more sensitive to scale.

It is worth noting that the explanatory power of employment density is significantly lower (almost halved) for checkerboard grids than for administrative units. Therefore,

[16] See for instance Combes et al. (forthcoming).

Table 5.7 – Gross wages and density

Simple correlations

	Dependent Variable: Log of gross wage (pooled years)					
Zoning system	(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
Density	0.071^a (0.001)	0.070^a (0.002)	0.073^a (0.001)	0.050^a (0.002)	0.090^a (0.003)	0.099^a (0.006)
Time dummies	yes	yes	yes	yes	yes	yes
Obs.	6138	6118	1692	1638	378	396
R^2	0.468	0.237	0.729	0.376	0.762	0.549

Notes: (*i*) All variables in logarithms. (*ii*) Standard errors in brackets.
(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels, respectively.

boundaries which do not reflect administrative/economic realities do actually generate measurement errors, possibly in both the left-hand and right-hand side variables. However, the good news is that these errors seem to be largely randomly distributed: even though density loses explanatory power, the overall picture with respect to elasticity is one of stability. In line with the intuitions provided in 5.2.3, this corroborates the OLS consistency in the presence of random measurement errors and exogenous explanatory variables.

As a second step, we compare the two MAUP distortions to the changes induced by including skills controls (Section 5.5.2) and market potential (Section 5.5.3) into the wage equation.

5.5.2 Controlling for skills and experience

Our empirical analysis uses rich individual wage information from a large panel of workers followed across time and jobs. We are hence able to apply a sophisticated procedure to control for observed and unobserved individual skills, so as to check whether the greater productivity observed in dense areas is partly due to the spatial sorting of workers and whether the MAUP affects these magnitudes. In a first stage, we calculate individual wages net of individual skills and experience, as follows:

$$\log w_{it} = \theta_i + \nu_{j(i,t)} + X_{it}\beta + \varepsilon_{it}, \tag{5.3}$$

where w_{it} is the wage of worker i at date t. This is a function of θ_i, an individual fixed-effect capturing the impact of both time-invariant observed and unobserved skills, $\nu_{j(i,t)}$, an effect specific to the firm j where i is employed at date t, and X_{it} a set of controls for worker's i experience at date t (age, age-squared, and number of previous jobs interacted with gender). Based on the estimates provided in Abowd, Creecy, and Kramarz (2002), and following Combes et al. (forthcoming), we define a wage net of any individual observed and unobserved skills and experience effects, $\left(w_{it} - \hat{\theta}_i - X_{it}\hat{\beta}\right)$. We then compute the average of this net wage over all individuals living in the same area a, at date t (hereafter net wage). This yields a measure of local labor productivity purged of individual skills and experience. We proceed by regressing net wages on employment density. The results are shown in table 5.8.

Table 5.8 – Net wages and density

Simple correlations

	Dependent Variable: Log of net wages (pooled years)					
Zoning system	(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
Density	0.033^a	0.028^a	0.029^a	0.023^a	0.048^a	0.052^a
	(0.001)	(0.001)	(0.001)	(0.002)	(0.003)	(0.004)
Time dummies	yes	yes	yes	yes	yes	yes
Obs.	6138	6118	1692	1638	378	396
R^2	0.220	0.098	0.338	0.238	0.619	0.570

Notes: (*i*) All variables in logarithms. (*ii*) Standard errors in brackets.
(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels, respectively.

The elasticity of net wages with respect to employment density is half of that for gross wages. Hence, the specification issue induces a difference in coefficient of an order of magnitude greater than that due to the MAUP. We therefore reach the same conclusion as for the analysis of spatial concentration: differences due to the size and shape of spatial units are small compared to the upward bias induced by the omission of workers' skills and experience in the wage equation, especially when data are not aggregated at a too large scale. Moreover, shape and size distortions are slightly attenuated in many cases (between DE and MS, and RE and LS, for instance), once these controls are included.

5.5.3 Market potential as a new control

Not only local density and skill composition affect labor performance, but so does the proximity to large economic centers outside the area. A major drawback of the above wage specifications is that there are no controls for the relative position of the area within the whole economy. For instance, wage equations derived from fully-specified economic geography models, such as Redding and Venables (2004) and Hanson (2005), account for spatial proximity via structural demand and supply access variables. It is beyond the scope of this paper to replicate such a sophisticated and difficult to implement approach. Here we only include, as well as density, a Harris (1954) market potential variable based on the employment accessible from any given area, divided by the distance necessary to reach them[17]:

$$\text{Market Potential} = \sum_{a' \neq a} \frac{Y_{a'}}{Dist_{a,a'}}, \tag{5.4}$$

where $Y_{a'}$ is employment in area a and $Dist_{a,a'}$, the great-circle distance between the centroids of areas a and a'. The results for gross and net wages are listed in tables 5.9

[17]The literature shows that this atheoretic market potential often has an explanatory power similar as the one of structural market potential.

and 5.10 respectively.

Table 5.9 – The spatial determinants of gross wages

	Dependent Variable: Log of gross wage (pooled years)					
Zoning system	(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
Density	0.055^a	0.065^a	0.059^a	0.050^a	0.090^a	0.098^a
	(0.001)	(0.002)	(0.002)	(0.002)	(0.003)	(0.006)
Market Potential	0.100^a	0.099^a	0.062^a	0.079^a	0.024^b	-0.009
	(0.004)	(0.008)	(0.005)	(0.008)	(0.011)	(0.020)
Obs.	6138	6118	1692	1638	378	396
R^2	0.521	0.256	0.753	0.411	0.765	0.549

Notes: (*i*) All variables in logarithms. (*ii*) Standard errors in brackets.
(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels, respectively.

Once market potential is accounted for, the impact of density on gross wage is attenuated. This is even more salient for low-scale and administrative zoning systems. The elasticity of gross wages to market potential is slightly stronger for medium squares than for their administrative counterparts, Départements. This is consistent with the intuition that cross-boundary discrepancies should be more salient for grid units that were not designed to minimize them in the first place.

Regarding the size issue, the impact of market potential monotonically decreases with the aggregation scale (for both the administrative and grid zoning systems). As for density, size distortions are more prevalent for RE or LS, and market potential becomes either insignificant or even negative. This is due to an important loss of information in the aggregation process, that we detail below.

Table 5.10 – The spatial determinants of net wages

	Dependent Variable: Log of net wage (pooled years)					
Zoning system	(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
Density	0.027^a	0.026^a	0.021^a	0.023^a	0.048^a	0.052^a
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)
Market Potential	0.037^a	0.043^a	0.036^a	0.044^a	0.023^b	-0.0002
	(0.004)	(0.007)	(0.006)	(0.007)	(0.01)	(0.012)
Obs.	6138	6118	1692	1638	378	396
R^2	0.232	0.104	0.354	0.256	0.624	0.570

Notes: (*i*) All variables in logarithms. (*ii*) Standard errors in brackets.
(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels, respectively.

In table 5.10 where skill controls are accounted for, shape and size alter only slightly the estimates at the lowest scales. It confirms our previous result that specification is of primary concern when working with small spatial units. Differences due to size and shape are much less pronounced than those resulting from a change in specification. For instance, the elasticity of density is only 0.027 at the small-unit levels, once skills and market potential are controlled for, while the baseline estimates were about 0.07. Similar conclusions are reached for the market potential elasticities, with slightly larger differences at the largest scales (RE and LS). To gain further insights on the underpinnings of such large-scale discrepancies, we turn to an alternative definition of market potential.

5.5.4 An alternative definition of market potential

If we use the average of municipality-level market potentials instead of the aggregate market potential, we obtain the results reported in tables 5.11 and 5.12.

Table 5.11 – The spatial determinants of gross wages:
Municipality-level market potential

	Dependent Variable: Log of gross wage (pooled years)					
Zoning system	(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
Density	0.050^a	0.061^a	0.051^a	0.038^a	0.063^a	0.069^a
	(0.001)	(0.002)	(0.002)	(0.002)	(0.004)	(0.006)
Market Potential	0.101^a	0.094^a	0.077^a	0.120^a	0.091^a	0.125^a
	(0.004)	(0.008)	(0.005)	(0.007)	(0.01)	(0.014)
Obs.	6138	6118	1692	1638	378	396
R^2	0.520	0.254	0.761	0.464	0.808	0.624

Notes: (*i*) All variables in logarithms. (*ii*) Standard errors in brackets.
(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels, respectively.

In this second set-up, the aggregation process conserves more information and, as expected, the elasticity of market potential is less sensitive to changes in the shape and size of units, and even less at the largest scales. Interestingly, the MAUP is also less salient regarding employment density, and the explanatory power of the model increases, in comparison with tables 5.9 and 5.10.

Table 5.12 – The spatial determinants of net wages:
Municipality-level market potential

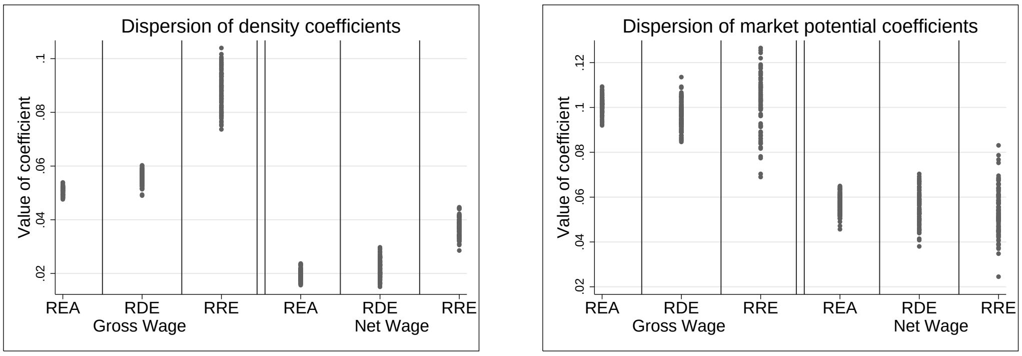
	Dependent Variable: Log of net wage (pooled years)					
Zoning system	(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
Density	0.025^a	0.024^a	0.017^a	0.016^a	0.032^a	0.038^a
	(0.001)	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)
Market Potential	0.038^a	0.044^a	0.042^a	0.063^a	0.056^a	0.059^a
	(0.004)	(0.006)	(0.006)	(0.007)	(0.009)	(0.009)
Obs.	6138	6118	1692	1638	378	396
R^2	0.232	0.105	0.357	0.278	0.652	0.611

Notes: (*i*) All variables in logarithms. (*ii*) Standard errors in brackets.
(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels, respectively.

As for net wages (see table 5.12), the coefficients of both density and market potential are more than halved compared to gross wages, whereas shape and size are clearly not big issues.

Figure 5.7, that displays the density and market potential estimates drawn from the three partly random zoning systems, provides further support to this conclusion. For a given size, the dispersion of estimates is much lower than that induced by a shift of specification, which confirms the absence of shape effects. Once again, the only significant difference due to size regards density for the largest units. Even so, this distortion almost vanishes in the best specification (net wages), as do the differences in the impact of market potential. These conclusions clearly echo the findings of Amrhein and Flowerdew (1992) and suggest that a good specification is actually an efficient way to circumvent the MAUP.

Figure 5.7 – The size- and shape-dependency of wage determinants



Note: (REA): Random employment areas, (RDE): Random Départements, (RE): Random Régions.

In line with the simulations provided in section 5.2.3, the loss of information incurred when variables are aggregated is the primary source of the MAUP. It can be mitigated (but never completely eliminated) when the process of aggregation is of the average-type and when the raw information is not too much heterogeneous *within-unit*, which is the case for spatially autocorrelated data at small scales. If so, the MAUP is of secondary concern compared to modeling issues.[18]

5.6 Gravity equations

So far, we have investigated MAUP distortions for aggregations processes that are of the average-type only. We now turn to gravity regressions that need both averaged and summed information.

[18]One important concern is not tackled here. In the above wage-density analysis, we inevitably face the major difficulty that causality could run both ways since the worker's location is also determined by their earnings anticipations. We leave this issue aside, as it has already been extensively discussed in the literature, and is orthogonal to the MAUP.

5.6.1 Basic gravity

The gravity model has been widely used to investigate the determinants of trade. A basic specification explains the trade flow $F_{aa'}$, originating from area a and shipped to area a', by various proxies for the proximity between a and a'. These include the great-circle distance between the centroids of a and a', $Dist_{aa'}$ and, often, a dummy variable stating whether the areas are contiguous, $Contig_{aa'}$.[19] Finally, the "border effect" (see McCallum, 1995) is captured by a dummy variable for within-area flows, $Within_{a=a'}$. As a first step, we estimate the following two-way fixed-effect specification:

$$\ln\left(F_{aa'}\right) = \theta_a + \theta_{a'} - \rho \ln\left(Dist_{aa'}\right) + \phi Contig_{aa'} + \psi Within_{a=a'} + \varepsilon_{aa'}, \quad (5.5)$$

where θ_a and θ'_a are destination and origin fixed effects, respectively, and $\varepsilon_{aa'}$ is an error term. This fixed-effect approach has the attractive property of being structurally compatible with many trade models (based on comparative advantage as well as imperfect competition).[20]

Table 5.13 – Basic gravity
Aggregate distance

	Dependent Variable: log of positive flows (Year 1996)					
Zoning system	(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
Distance	-0.996^a	-1.175^a	-1.608^a	-1.912^a	-1.602^a	-1.900^a
	(0.022)	(0.024)	(0.056)	(0.048)	(0.075)	(0.113)
Within	1.738^a	1.040^a	1.395^a	0.221	1.460^a	0.445^b
	(0.063)	(0.066)	(0.111)	(0.135)	(0.151)	(0.211)
Contiguity	0.967^a	1.093^a	0.959^a	1.044^a	0.728^a	0.895^a
	(0.041)	(0.044)	(0.063)	(0.077)	(0.087)	(0.118)
Obs.	24849	22189	6600	5069	441	443
R^2	0.516	0.541	0.706	0.752	0.941	0.928

Notes: (*i*) All variables in logarithms. (*ii*) Standard errors in brackets.
(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels, respectively.

Table 5.13 reports on the related estimates under both the administrative and grid zoning systems. The great-circle distance elasticity is systematically larger for grid than for administrative zoning systems, at a given scale. The shape effect on distance increases with the scale of aggregation. Contiguity is less affected by shape. Again, size effects are slightly more salient at the largest scales, especially when moving from the EA-SS to either the DE-MS or RE-LS zoning systems. The magnitude of the distance effect (in absolute value) increases with size (for the administrative and grid zoning systems). The border effect is always lower for grid zoning systems, which is further evidence of the economic consistency of administrative units.

[19] A for grid zoning systems, we assume that two units are contiguous if they share a common edge.
[20] See Feenstra (2003).

If we use the average of inter-municipality distance instead of aggregate distance (see table 5.14), results remain virtually the same, but the border effect is magnified.

Table 5.14 – Basic gravity
Municipality-level distance

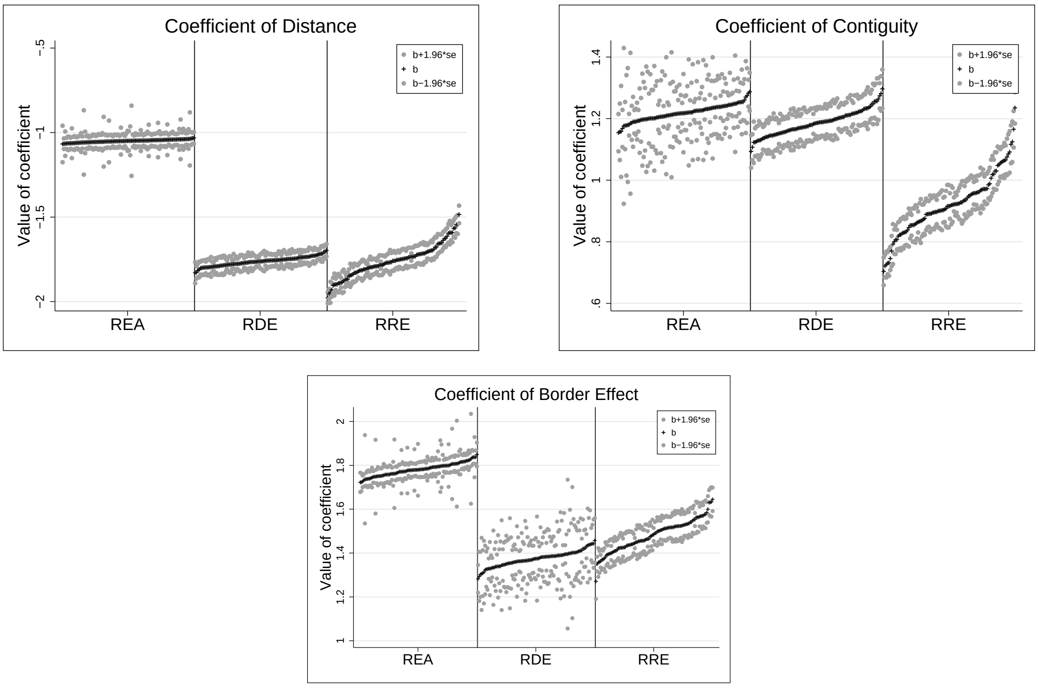
	Dependent Variable: log of positive flows (Year 1996)					
Zoning system	(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
Weighted Distance	-1.009^a	-1.182^a	-1.645^a	-1.909^a	-1.710^a	-1.968^a
	(0.022)	(0.024)	(0.058)	(0.047)	(0.088)	(0.096)
Within	2.139^a	1.547^a	1.938^a	1.138^a	1.900^a	1.395^a
	(0.058)	(0.056)	(0.099)	(0.097)	(0.146)	(0.21)
Contiguity	1.031^a	1.139^a	1.020^a	1.058^a	0.768^a	0.863^a
	(0.04)	(0.044)	(0.062)	(0.069)	(0.082)	(0.094)
Obs.	24849	22189	6600	5069	441	443
R^2	0.517	0.544	0.709	0.757	0.942	0.933

Notes: (*i*) All variables in logarithms. (*ii*) Standard errors in brackets.
(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels, respectively.

In sharp contrast with market potential in wage equations, an alternative measure of distance does not alleviate the MAUP. Gravity regressions are hence more sensitive to the MAUP. The rationale is found in the simulations depicted in section 5.2.3. The dependent variable, trade flows, is summed over units, whereas the explanatory variable, distance, is averaged. The process of aggregation shifts to the right the distribution of the former and raises its dispersion (which finds support in table 5.1). By way of contrast, since distance is a highly autocorrelated averaged variable, it is less sensitive to aggregation. The rise (in absolute value) of the distance coefficient reflects the need to reconciliate an increasing dispersion of trade flows with a stable support of the distance distribution.

Figure 5.8 illustrates the way in which both size and shape affect the values and standard errors of estimates from partly random zoning systems. Dark dots in the top-left figure stand for the elasticity of distance (and for contiguity and border effects in the top-right and bottom figures, respectively). The 95% confidence interval is shown by the surrounding lighter dots. Random zoning systems are ranked by increasing estimated values. For all three proximity measures, we find that the variability in estimates raises with scale (as reflected by the increasing slope of dark curves), suggesting more shape-dependency in larger zoning systems. Nonetheless, this variability is of lower magnitude than the differences due to moving from one scale to another (from REA to RDE or RRE, regarding distance and border effects). The shape-dependency of larger zoning systems (especially RRE) is due to two joint phenomena. First, coefficient estimation is more likely to suffer from finite-sample bias for larger (and hence less numerous) units. Second, the random process of aggregation is likely to produce more distinct zoning systems when data are aggregated over larger units.

Figure 5.8 – The size- and shape-dependency of the impact of spatial proximity on trade



Notes: (*i*) The coefficients (b) have to be greater (in absolute value) than 1.96 times the standard error (se) to enter into the 95% confidence interval. (*ii*) (REA): Random employment areas, (RDE): Random Départements, (RRE): Random Régions.

5.6.2 Augmented Gravity

Barriers to trade do not only concern proximity. Other trade frictions result from costs unrelated to distance (such as trade policy, exchange-rate volatility, delivery times, and inventory or regulation costs), and from more subtle frictions due to the need to acquire information on remote trading partners or to enforce contracts, as emphasized by Rauch (2001). To tackle these, the literature extends the basic gravity model by making trade costs depend not only on spatial proximity but also on cultural and informational proximity. For instance Wagner et al. (2002) report that migrations between two countries enhance their bilateral trade by around 50%. To evaluate the trade-creating impact of social and business networks within countries, Combes et al. (2005) estimate:

$$\begin{aligned}\ln(F_{aa'}) &= \theta_a + \theta_{a'} - \rho \ln(Dist_{aa'}) + \phi Contig_{aa'} + \psi Within_{a=a'} \\ &\quad + \alpha \ln(1 + Mig_{aa'}) + \beta \ln(1 + Mig_{a'a}) + \gamma \ln(1 + Plant_{aa'}) + \varepsilon_{aa'},\end{aligned} \tag{5.6}$$

where $Dist_{aa'}$ is municipality-level distance,[21] $Mig_{aa'}$ is the number of people born in area a' and working in area a, called (relative to area a) immigrants, $Mig_{a'a}$ are analogously emigrants, and $Plant_{aa'}$ is the number of financial connections between plants belonging to the same business group (see 5.8).

Table 5.15 – Augmented Gravity

	Dependent Variable: log of positive flows (year 1996, Municipality-level distance)					
Zoning system	(EA)	(SS)	(DE)	(MS)	(RE)	(LS)
Weighted Distance	-0.616^a	-0.698^a	-1.231^a	-1.294^a	-1.291^a	-1.340^a
	(0.023)	(0.027)	(0.062)	(0.061)	(0.103)	(0.102)
Within	1.201^a	0.925^a	0.8^a	0.338^a	0.517^a	0.436^c
	(0.064)	(0.06)	(0.126)	(0.095)	(0.171)	(0.241)
Contiguity	0.315^a	0.403^a	0.366^a	0.317^a	0.296^a	0.425^a
	(0.049)	(0.049)	(0.068)	(0.072)	(0.072)	(0.119)
Emigrants	0.228^a	0.226^a	0.237^a	0.244^a	0.281^a	0.246^b
	(0.014)	(0.013)	(0.028)	(0.034)	(0.088)	(0.104)
Immigrants	0.241^a	0.256^a	0.209^a	0.286^a	0.257^a	0.268^b
	(0.014)	(0.015)	(0.037)	(0.035)	(0.086)	(0.134)
Business networks	0.043^a	0.013	0.24^a	-0.021	0.225	0.646^a
	(0.016)	(0.019)	(0.072)	(0.064)	(0.173)	(0.161)
Obs.	24849	22189	6600	5069	441	443
R^2	0.538	0.568	0.723	0.772	0.953	0.945

Notes: (*i*) All variables in logarithms. (*ii*) Standard errors in brackets.
(*iii*) ^{a}, ^{b}, ^{c}: Significant at the 1%, 5% and 10% levels, respectively.

It can readily be seen from table 5.15 that, controlling for networks reduces the distance elasticity by about one-third, whereas the contiguity effect is three to four times smaller. The border effect is reduced even further, and disappears completely at the RE-LS scales.

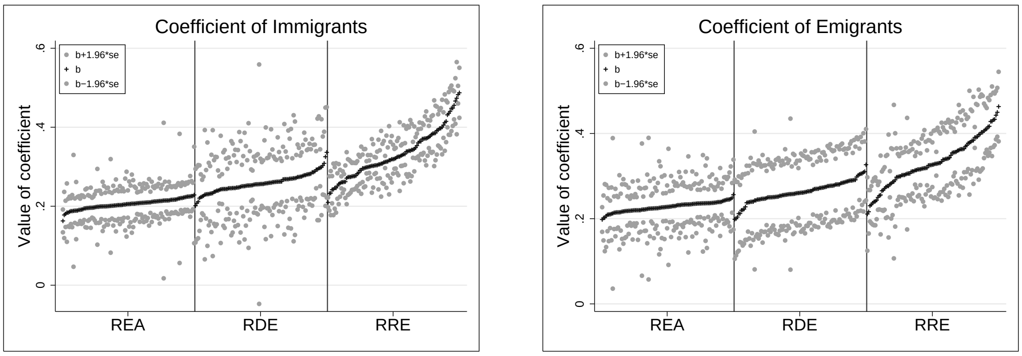
[21] Results are virtually unchanged with the alternative measure of distance, i.e. aggregate distance.

The MAUP distortions are subsequently far larger than those observed in table 5.13 and 5.14.

It is worth noting that the trade-creating effect of migrants is robust to the shift of zoning system, in terms of both size and shape. Migrant and business network variables are indeed summed from one scale to another, and this aggregation process increases both their mean and dispersion. Their elasticity is not very sensitive to the MAUP because the dependent variable, trade, is aggregated under the same summation process. By way of contrast, even though the trade-creating impact of business networks increases slightly with the scale of administrative units, it is no longer statistically significant for grid zoning systems.

Figure 5.9 displays the estimated immigrant and emigrant coefficients in the same way as in figure 5.8. Both groups of estimates monotonically increase with the level of aggregation.

Figure 5.9 – The size- and shape-dependency of the trade-creating impact of migrants



Notes: (*i*) The coefficients (*b*) have to be greater (in absolute value) than 1.96 times the standard error (se) to enter into the 95% confidence interval. (*ii*) (REA): Random employment areas, (RDE): Random Départements, (RRE): Random Régions.

We therefore continue to find that size matters more than shape. Moreover, the magnitude of this distortion is definitely larger than in our previous exercises. The explanation is that gravity regressions involve variables aggregated under different processes. Since the MAUP is fundamentally linked to whether the distribution of variables is preserved, it jeopardizes gravity estimations more than wage equations. Still, MAUP distortions remain of smaller magnitude than mis-specification biases.

5.7 Conclusion

The overall picture is fairly clear. The use of different specifications to assess spatial concentration, agglomeration economies, and trade determinants produces substantial variation in the estimated coefficients. In most cases, theory provides a clear explanation of such variations. Although the size effect of the MAUP might still be important, especially at large scales, it is of second-order compared to specification at lower scales. Shape distortions remain of only third-order concern. On the other hand, when zoning systems are specifically designed to address local questions, as is the case for French employment areas, we definitely argue that they should be used. Those who are left with other administrative units should not worry too much however, as long as the aggregation scale is not too large. We therefore urge researchers to pay attention in priority to choosing the relevant specification for the question they want to tackle.

We also want to draw attention on the fact that the aggregation process conditions the magnitude of the MAUP distortions. If these distortions are negligible when both the dependent and explanatory variables are averaged, they are clearly more jeopardizing when the aggregation processes are not consistent on both sides of the regression, and even more that we work with large-scale spatial units. For instance, the MAUP could be of greater concern with U.S. data aggregated at the State level.

We do not of course claim that the various specifications used in this paper are actually the best. They are simply those frequently found in the economic geography literature. Many other empirical questions can be considered. We focus on three simple exercises because they are quite different in spirit, and cover a wide range of estimations. This makes us fairly confident that our conclusions are robust to other exercises, even though this remains to be shown.

Finally, the French economical and institutional design may be particularly well-designed to minimize MAUP problems. For instance, the division of France into Départements, was adopted simultaneously with the first French constitution in 1790 to replace the old "provinces", which more or less represented dioceses. These latter exhibited significant variation in tax systems, population and land areas, and the new division aimed to create more "regular" spatial units under a common central legislation and administration. Their size was chosen so that individuals from any point in the Département could make the round trip by horse to the capital city in no more than two days, which translated into a radius of 30-40 km. Hence, it might well be that the French administrative zoning systems are less sensitive to the MAUP by definition. We therefore encourage researchers to replicate the exercises carried out here in the context of other countries.

5.8 Appendix to chapter 5: Data

Economic variables for all zoning systems are obtained by aggregating information over the 36,247 French municipalities ("communes").

First, over the 1976-1996 period, the composition in terms of establishments (employment size, and number of establishments) and workers (year and place of birth, age, gender, occupation, and wage, among others) is available at the 4-digit industrial level. The data come from the INSEE survey "*Déclaration Annuelle de Données Sociales*" (DADS), which collects matched employer-employee information in France. Our analysis builds on a panel extract covering people born in October of all even-numbered years, excluding civil servants, which is a representative $1/24^{th}$ of the French population. No survey was carried out in 1981, 1983 or 1990, producing a final sample of over 12.3 million plant - individual year observations, which are then re-aggregated by spatial unit, year (18 points), and industry (98 two-digit sectors covering both manufacturing and services).[22] As the key parameter of the sampling process is the date of birth, there is no obvious reason to believe that the sample is geographically biased.

For 1996, the above data are matched with information on the trade volumes shipped by road, both within and between municipalities, which we aggregate into different larger zoning systems. The data comes from the French Ministry of Transport, which annually surveys a stratified random sample of trucks.

Regarding social and business networks, we compute migrant stocks based on the number of natives from one area who moved to work in another area.[23] Business networks are captured via the number of financial connections between plants belonging to the same business group. For each business group, we count the number of plants located in each area. We then compute for each pair of areas the sum over all business groups of the product of the two counts. The data source here is the INSEE survey "*LIaisons FInancières*" (LIFI), which defines a business group as the set of all firms controlled either directly or indirectly (over 50%) by the same parent firm, which is itself not controlled by any other firm.[24]

[22]As in Abowd et al. (2002), part-timers are retained and outliers (over five standard errors above and below the mean) are dropped. The selection of industries and the removal of sampling errors at the smallest scale follows Combes et al. (forthcoming).

[23]This figure is also calculated using the DADS survey.

[24]See Combes et al. (2005) for more details on the network variables.

Bibliography

Abowd, J. M., Creecy, R. H., Kramarz, F., 2002. Computing person and firm effects using linked longitudinal employer-employee data. Technical Paper 2002-06, Longitudinal Employer-Household Dynamics, Center for Economic Studies, U.S. Census Bureau.

Ackerberg, D., Benkard, C. L., Berry, S., Pakes, A., 2007. Econometric tools for analyzing market outcomes. In: Heckman, J. J., Leamer, E. (Eds.), Handbook of Econometrics. Vol. 6A. Elsevier Science Publisher, Ch. 63, pp. 4171–4276.

Amrhein, C. G., 1995. Searching for the elusive aggregation effect: Evidence from statistical simulations. Environment and Planning A 27, 105–119.

Amrhein, C. G., Flowerdew, R., 1992. The effect of data aggregation on a poisson regression model of Canadian migration. Environment and Planning A 24, 1381–1391.

Amrhein, C. G., Reynolds, H., 1997. Using the Getis statistic to explore aggregation effects in metropolitan Toronto census data. The Canadian Geographer 41 (2), 137–149.

Anderson, J. E., Marcouiller, D., 2002. Insecurity and the pattern of trade: An empirical investigation. The Review of Economics and Statistics 84 (2), 342–352.

Anderson, J. E., VanWincoop, E., 2003. Gravity with gravitas: A solution to the border puzzle. American Economic Review 93 (1), 170–192.

Anderson, J. E., VanWincoop, E., 2004. Trade costs. Journal of Economic Literature 42 (3), 691–751.

Arbia, G., 1989. Spatial Data Configuration in Statistical Analysis of Regional Economic and Related Problems. Kluwer.

Arbia, G., Espa, G., Giuliani, D., Mazzitelli, A., 2009. Clusters of firms in space and time. Working Paper 0902, Department of Economics, University of Trento, Italia.

Arbia, G., Espa, G., Quah, D., 2008. A class of spatial econometric methods in the empirical analysis of clusters of firms in the space. Empirical Economics 34 (1), 81–103.

Arzaghi, M., Henderson, J. V., 2008. Networking off Madison avenue. Review of Economic Studies 75 (4), 1011–1038.

Aubert, P., Crépon, B., 2003. La productivité des salariés: une tentative d'estimation. Economie et statistique 368, 95–119.

Baier, S. L., Bergstrand, J. H., 2009. Bonus vetus OLS: A simple method for approximating international trade-cost effects using the gravity equation. Journal of International Economics 77 (1), 77–85.

Bandyopadhyay, S., Coughlin, C. C., Wall, H. J., 2008. Ethnic networks and US exports. Review of International Economics 16 (1), 199–213.

Barbesol, Y., Briant, A., 2009. Économies d'agglomération et productivité des entreprises : estimation sur données individuelles françaises. Économie et Statistique 419-420, 31–54.

Barrios, S., Bertinelli, L., Strobl, E., Teixeira, A.-C., 2005. The dynamics of agglomeration: Evidence from Ireland and Portugal. Journal of Urban Economics 57 (1), 170–188.

Bartel, A. P., 1989. Where do the new US immigrants live? Journal of Labor Economics 7 (4), 371–91.

Baum, C. F., Schaffer, M. E., Stillman, S., 2003. Instrumental variables and GMM: Estimation and testing. Stata Journal 3 (1), 1–31.

Behrens, K., Ertur, C., Koch, W., 2007. Dual gravity: Using spatial econometrics to control for multilateral resistance, CORE Discussion Paper 2007/59.

Bénassy-Quéré, A., Coupet, M., Mayer, T., 2007. Institutional determinants of foreign direct investment. The World Economy 30 (5), 764–782.

Berkowitz, D., Moenius, J., Pistor, K., 2006. Trade, law, and product complexity. The Review of Economics and Statistics 88 (2), 363–373.

Bound, J., Jaeger, D. A., Baker, R. M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. Journal of the American Statistical Association 90, 443–450.

Briant, A., Combes, P.-P., Lafourcade, M., 2010. Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? Journal of Urban Economics forthcoming.

Buchinsky, M., 1998. Recent advances in quantile regression models: A practical guideline for empirical research. The Journal of Human Resources 33 (1), 88–126.

Burnod, G., Chenu, A., 2001. Employés qualifiés et non-qualifiés: une proposition d'aménagement de nomenclature des catégories socio-professsionnelles. Travail et Emploi 86.

Cameron, C., Trivedi, P., 2005. Microeconometrics: Methods and Applications. Cambridge Uinversity Press.

Casella, A., Rauch, J., 2003. Overcoming informational barriers to international resource allocation: Prices and group ties. Economic Journal 113 (484), 21–42.

Ciccone, A., 2002. Agglomeration effects in Europe. European Economic Review 46, 213–227.

Ciccone, A., Hall, R., 1996. Productivity and the density of economic activity. American Economic Review 86 (1), 54–70.

Cingano, F., Schivardi, F., 2004. Identifiying the sources of local productivity growth. Journal of European Economic Association 2 (4), 720–742.

Combes, P.-P., 2000. Economic structure and local growth: France, 1984-1993. Journal of Urban Economics 47 (3), 329–355.

Combes, P.-P., Duranton, G., Gobillon, L., 2008a. Spatial wage disparities: Sorting matters! Journal of Urban Economics 63 (2), 723–742.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., Roux, S., 2009. The productivity advantages of large markets: Distinguishing agglomeration from firm selection. C.E.P.R. Discussion Papers 7191, Centre for Economic Policy Research.

Combes, P.-P., Duranton, G., Gobillon, L., Roux, S., forthcoming. Estimating agglomeration economies with history, geology and worker effects. In: Glaeser, E. L. (Ed.), The Economics of Agglomeration. University of Chicago Press.

Combes, P.-P., Lafourcade, M., 2005. Transport costs: measures, determinants, and regional policy implications for france. Journal of Economic Geography 5 (3), 319–349.

Combes, P.-P., Lafourcade, M., Mayer, T., 2005. The trade-creating effects of business and social networks: Evidence from France. Journal of International Economics 66 (1), 1–29.

Combes, P.-P., Magnac, T., Robin, J.-M., 2004. The dynamics of local employment in France. Journal of Urban Economics 56 (2), 217–243.

Combes, P.-P., Mayer, T., Thisse, J.-F., 2008b. Economic Geography. Princeton University Press.

Combes, P.-P., Overman, H. G., 2004. The spatial distribution of economic activities in the European Union. In: Henderson, J. V., Thisse, J. F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4: Cities and Geography. Elsevier, Ch. 64, pp. 2845–2909.

Crépon, B., Deniau, N., Pérez-Duarte, S., 2002. Wages, productivity, and worker characteristics: A French perspective, mimeo.

Crépon, B., Duhautois, R., 2003. Ralentissement de la productivité et réallocations d'emplois: deux régimes de croissance. Économie et Statistique 367, 69–82.

de Groot, H. L. F., Linders, G.-J., Rietveld, P., Subramanian, U., 2004. The institutional determinants of bilateral trade patterns. Kyklos 57 (1), 103–123.

Devereux, M. P., Griffith, R., Simpson, H., 2004. The geographic distribution of production activity in the U.K. Regional Science and Urban Economics 34 (5), 533–564.

Diggle, P. J., 2003. Statistical analysis of spatial point patterns. Oxford University Press.

Dumais, G., Ellison, G., Glaeser, E. L., 2002. Geographic concentration as a dynamic process. The Review of Economics and Statistics 84 (2), 193–204.

Dunlevy, J., 2006. The influence of corruption and language on the protrade effect of immigrants: Evidence from the American States. The Review of Economics and Statistics 88 (1), 182–186.

Duranton, G., Martin, P., Mayer, T., Mayneris, F., 2008. Les pôles de compétitivité: Que peut-on en attendre ? Centre pour la Recherche Économique et ses Applications - Éditions de l'École Normale Supérieure (CEPREMAP/ENS).

Duranton, G., Overman, H. G., 2005. Testing for localization using micro-geographic data. Review of Economic Studies 72 (4), 1077–1106.

Duranton, G., Overman, H. G., 2008. Exploring the detailed location patterns of U.K. manufacturing industries using microgeographic data. Journal of Regional Science 48 (1), 213–243.

Duranton, G., Puga, D., 2001. Nursery cities: Urban diversity, process innovation and the life-cycle of product. American Economic Review 91 (5), 1454–1477.

Duranton, G., Puga, D., 2004. Micro-foundations of urban agglomeration economies. In: Henderson, J. V., Thisse, J. F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4: Cities and Geography. Elsevier, Ch. 48, pp. 2063–2117.

Eaton, J., Tamura, A., 1994. Bilateralism and regionalism in Japanese and U.S. trade and direct foreign investment patterns. Journal of the Japanese and International Economies 8, 478–510.

Ellison, G., Glaeser, E. L., 1997. Geographic concentration in U.S. manufacturing industries: A dartboard approach. Journal of Political Economy 105 (5), 889–927.

Ellison, G., Glaeser, E. L., 1999. The geographic concentration of industry: Does natural advantage explain agglomeration? American Economic Review 89 (2), 311–316.

Ellison, G., Glaeser, E. L., Kerr, W., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. American Economic Review forthcoming.

ESPON, 2006. The modifiable areas unit problem. Tech. rep., European Spatial Planning Observation Network.

Feenstra, 2003. Advanced International Trade: Theory and Evidence. Princeton University Press.

Feenstra, R. C., 1996. U. S. imports, 1972—1994: Data and concordance, Working Paper 5515, National Bureau of Economic Research.

Foster, L., Haltiwanger, J., Syverson, C., 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? American Economic Review 98 (1), 394–425.

Fotheringham, A. S., Wong, D. W. S., 1991. The modifiable areal unit problem in multivariate statistical analysis. Environment and Planning A 23, 1025–1044.

Fujita, M., Mori, T., Henderson, J. V., Kanemoto, Y., 2004. Spatial distribution of economic activities in Japan and China. In: Henderson, J. V., Thisse, J. F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4: Cities and Geography. Elsevier, Ch. 65, pp. 2911–2977.

Fuss, M., McFadden, D., 1978. Production Economics: a Dual Approach to Theory and Applications. North Holland.

Gehlke, C., Biehl, K., 1934. Certain effects on grouping upon the size of the correlation coefficient in census tract material. Journal of the American Statistical Association 29 (185), 169–170.

Girma, S., Yu, Z., 2002. The link between immigration and trade: Evidence from the United Kingdom. Weltwirtschaftliches Archiv 138 (1), 115–130.

Glaeser, E. L., 2008. Cities, Agglomeration and Spatial Equilibrium. Oxford University Press.

Glaeser, E. L., Kallal, H., Scheinkman, J., Schleifer, A., 1992. Growth in cities. Journal of Political Economy 100 (6), 1126–1152.

Glaeser, E. L., Maré, D., 2001. Cities and skills. Journal of Labor Economics 19 (2), 316–342.

Gould, D., 1994. Immigrant links to the home country: Empirical implications for U.S. bilateral trade flows. The Review of Economics and Statistics 76 (2), 302–316.

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: Applications to Poisson models. Econometrica 52 (3), 701–20.

Griliches, Z., Mairesse, J., 1995. Production functions: The search for identification. NBER Working Papers 5067, National Bureau of Economic Research, Inc.

Guillain, R., Le Gallo, J., 2007. Agglomeration and dispersion of economic activities in Paris and its surroundings: An exploratory spatial data analysis, mimeo.

Hanson, G., 2005. Market potential, increasing returns, and geographic concentration. Journal of International Economics 67, 1–24.

Harris, C., 1954. The market as a factor in the localization of industry in the United States. Annals of the Association of American Geographers 44, 315–348.

Head, K., Mayer, T., 2004. The empirics of agglomeration and trade. In: Henderson, J. V., Thisse, J. F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4: Cities and Geography. Elsevier, Ch. 59, pp. 2609–2669.

Head, K., Mayer, T., 2006. Regional wage and employment responses to market potential in the EU. Regional Science and Urban Economics 36 (5), 573–594.

Head, K., Mayer, T., Ries, J., 2008. The erosion of colonial trade linkages after independence. Discussion Paper 6951, CEPR.

Head, K., Mayer, T., Ries, J., 2009. How remote is the offshoring threat? European Economic Review 53 (4), 429–444.

Head, K., Ries, J., 1998. Immigration and trade creation: Econometric evidence from Canada. Canadian Journal of Economics 31 (1), 47–62.

Hellerstein, J., Neumark, D., Troske, K., 1999. Wage, productivity and workers characteristics: Evidence from the plant-level production functions and wage equations. Journal of Labor Economics 13 (3), 409–446.

Helpman, E., Melitz, M., Rubinstein, Y., 2008. Estimating trade flows: Trading partners and trading volumes. The Quarterly Journal of Economics 123 (2), 441–487.

Henderson, J. V., 1974. The sizes and types of cities. American Economic Review 64 (4), 640–56.

Henderson, J. V., 1986. Efficiency of resource usage and city size. Journal of Urban Economics 19 (1), 47–70.

Henderson, J. V., 1997. Externalities and industrial development. Journal of Urban Economics 42 (3), 449–470.

Henderson, J. V., 2003. Marshall's scale economies. Journal of Urban Economics 53 (1), 1–28.

Henderson, J. V., Kuncoro, A., Turner, M., 1995. Industrial development in cities. Journal of Political Economy 103 (5), 1067–1090.

Herander, M., Saavedra, L., 2005. Exports and the structure of immigrant-based networks: the role of geographic proximity. The Review of Economics and Statistics 87 (2), 323–335.

Holmes, T. J., Lee, S., forthcoming. Cities as six-by-six-mile squares: Zipf's law? In: Glaeser, E. L. (Ed.), The Economics of Agglomeration. University of Chicago Press.

Holmes, T. J., Stevens, J. J., 2002. Geographic concentration and establishment scale. The Review of Economics and Statistics 84 (4), 682–690.

Holmes, T. J., Stevens, J. J., 2004. Spatial distribution of economic activities in North America. In: Henderson, J. V., Thisse, J. F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4: Cities and Geography. Elsevier, Ch. 63, pp. 2797–2843.

Holt, D., Steel, D., Tranmer, M., Wrigley, N., 1996. Aggregation and ecological effects in geographically-based data. Geographical Analysis 28, 244–261.

Jacobs, J., 1969. The Economy of Cities. Random House, Inc.

Jayet, H., Bolle-Ukrayinchuk, N., 2007. The location of immigrants in France, mimeo.

Kaufmann, D., Kraay, A., Mastruzzi, M., 2007. Governance matters VI: Aggregate and individual governance indicators 1996-2006. Tech. rep., The World Bank.

Klier, T., McMillen, D., 2008. Evolving agglomeration in the U.S. auto supplier industry. Journal of Regional Science 48 (1), 245–267.

Koenker, R., Bassett, J. G., 1978. Regression quantiles. Econometrica 46, 33–50.

Kolko, J., 1999. Can I get some service here? Information technology, service industries and the future of cities, mimeo.

Kolko, J., forthcoming. Urbanization, agglomeration, and co-agglomeration of service industries. In: Glaeser, E. L. (Ed.), The Economics of Agglomeration. University of Chicago Press.

Krugman, P., 1991. Increasing returns and economic geography. Journal of Political Economy 99 (3), 483–99.

Levinsohn, J., Petrin, A., 2003. Estimating production functions using inputs to control for unobservables. Review of Economic Studies 70 (2), 317–342.

Lucas, R. J. E., July 1988. On the mechanics of economic development. Journal of Monetary Economics 22 (1), 3–42.

Lucas, R. J. E., April 2001. Externalities and cities. Review of Economic Dynamics 4 (2), 245–274.

Marcon, E., Puech, F., 2003. Evaluating the geographic concentration of industries using distance-based methods. Journal of Economic Geography 3(4), 409–428.

Marcon, E., Puech, F., 2007. Measures of the geographic concentration of industries: Improving distance-based methods, mimeo.

Marschak, J., Andrews, W., 1944. Random simultaneous equations and the theory of production. Econometrica 12, 143–205.

Marshall, A., 1890. Principles of Economics, 8th Edition. MacMillan and Co.
URL `http://www.econlib.org/library/Marshall/marP.html`

Martin, P., Mayer, T., Mayneris, F., 2008. Spatial concentration and firm-level productivity in France. C.E.P.R. Discussion Papers 6858, Centre for Economic Policy Research.

Maurel, F., Sédillot, B., 1999. A measure of geographic concentration in French manufacturing industries. Regional Science and Urban Economics 29 (5), 575–604.

McCallum, J., 1995. National borders matter: Canada-u.s. regional trade patterns. American Economic Review 85 (3), 615–23.

Melitz, M. J., 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. Econometrica 71 (6), 1695–1725.

Melitz, M. J., Ottaviano, G. I. P., 2008. Market size, trade, and productivity. Review of Economic Studies 75 (1), 295–316.

Melo, P. C., Graham, D. J., Noland, R. B., 2009. A meta-analysis of estimates of urban agglomeration economies. Regional Science and Urban Economics 39 (3), 332–342.

Menon, C., 2008. The bright side of MAUP: an inquiry on the determinants of industrial agglomeration in the United States. Working Paper 2008/29, Ca'Foscari University of Venice.

Millimet, D., Osang, T., 2007. Does state borders matter for U.S. intranational trade? The role of history and internal migration. Canadian Journal of Economics 40 (1), 93–126.

Moomaw, R. L., 1981. Productivity and city size? a critique of the evidence. The Quarterly Journal of Economics 96 (4), 675–88.

Moretti, E., 2004. Workers' education, spillovers, and productivity: Evidence from plant-level production functions. American Economic Review 94 (3), 656–690.

Mori, T., Nishikimi, K., Smith, T. E., 2005. A divergence statistic for industrial localization. The Review of Economics and Statistics 87 (4), 635–651.

Moulton, B. R., 1990. An illustration of a pitfall in estimating the effects of aggregate variables on micro unit. The Review of Economics and Statistics 72 (2), 334–338.

Munshi, K., 2003. Networks in the modern economy: Mexican migrants in the U.S. labor market. The Quarterly Journal of Economics 118 (2), 549–599.

Nakamura, R., 1985. Agglomeration economies in urban manufacturing industries: A case of Japanese cities. Journal of Urban Economics 17 (1), 108–124.

Nocke, V., 2006. A gap for me: Entrepreneurs and entry. Journal of the European Economic Association 4 (5), 929–956.

Okubo, T., Picard, P. M., Thisse, J.-F., 2008. The spatial selection of heterogeneous firms. Discussion Paper 6978, C.E.P.R.

Olley, G. S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64 (6), 1263–98.

Openshaw, S., Taylor, P., 1979. A million of so correlation coefficients: Three experiments on the modifiable areal unit problem. In: Wrigley, N. (Ed.), Statistical Applications in the Spatial Sciences. Pion London, pp. 127–144.

Ottaviano, G., Thisse, J.-F., 2004. Agglomeration and economic geography. In: Henderson, J. V., Thisse, J. F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4: Cities and geography. Elsevier, Ch. 58, pp. 2563–2608.

Pepper, J. V., 2002. Robust inferences from random clustered samples: an application using data from the panel study of income dynamics. Economics Letters 75 (3), 341–345.

Puga, D., 2009. The magnitude and causes of agglomeration economies. Working Paper 2009-09, Instituto Madrileño de Estudios Avanzados (IMDEA) Ciencias Sociales.

Ranjan, P., Lee, J. Y., 2007. Contract enforcement and international trade. Economics and Politics 19 (2), 191–218.

Rauch, J., 1993. Does history matter only when it matters little? the case of city-industry location. The Quarterly Journal of Economics 108 (3), 843–867.

Rauch, J., 1999. Networks versus markets in international trade. Journal of International Economics 48 (1), 7–35.

Rauch, J., 2001. Business and social networks in international trade. Journal of Economic Literature 39 (4), 1177–1203.

Rauch, J., Trindade, V., 2002. Ethnic chinese networks in international trade. The Review of Economics and Statistics 84 (1), 116–130.

Redding, S., Venables, A., 2004. Economic geography and international inequality. Journal of International Economics 62, 53–82.

Reynolds, H. D., 1998. The modifiable areal unit problem: Empirical analysis by statistical simulation. Ph.D. thesis, University of Toronto, Department of Geography.

Roback, J., 1982. Wages, rents, and the quality of life. Journal of Political Economy 90, 1257–1278.

Rosen, S., 1979. Wage-based indexes of urban quality of life. In: Miezkowski, P., Straszheim, M. R. (Eds.), Current Issues in Urban Economics. Baltimore: John Hopkins University Press.

Rosenthal, S. S., Strange, W. C., 2003. Geography, industrial organization, and agglomeration. The Review of Economics and Statistics 85 (2), 377–393.

Rosenthal, S. S., Strange, W. C., 2004. Evidence on the nature and sources of agglomeration economies. In: Henderson, J. V., Thisse, J. F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4: Cities and Geography. Elsevier, Ch. 49, pp. 2119–2171.

Rosenthal, S. S., Strange, W. C., 2008. The attenuation of human capital spillovers. Journal of Urban Economics 64, 373–389.

Rosenthal, S. S., Strange, W. C., forthcoming. Small establishments/big effects: Agglomeration, industrial organization and entrepreneurship. In: Glaeser, E. L. (Ed.), The Economics of Agglomeration. University of Chicago Press.

Santos Silva, J., Tenreyro, S., 2006. The log of gravity. The Review of Economics and Statistics 88 (4), 641–658.

Silverman, B., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall.

Staiger, D., Stock, J. H., 1997. Instrumental variables regression with weak instruments. Econometrica 65 (3), 557–586.

Starrett, D., 1978. Market allocations of location choice in a model with free mobility. Journal of Economic Theory 17 (1), 21–37.

Strange, W. C., 2009. Viewpoint: Agglomeration research in the age of disaggregation. Canadian Journal of Economics 42 (1), 1–27.

Thierry, X., 2004. Evolution de l'immigration en France et éléments de comparaison avec le Royaume-Uni. Population 5, 725–764.

Turrini, A., van Ypersele, T., 2006. Legal costs as barriers to trade. Discussion Paper 5751, CEPR.

Wagner, D., Head, K., Ries, J., 2002. Immigration and the trade of provinces. Scottish Journal of Political Economy 49 (5), 507–25.

Wallace, N., Walls, D., 2004. Agglomeration economies and the high-tech computer cluster, mimeo.

Windmeijer, F., 2006. GMM for panel count data models. Bristol Economics Discussion Papers 06/591, Department of Economics, University of Bristol, UK.

Wooldridge, J. M., 2003. Cluster-sample methods in applied econometrics. American Economic Review 93 (2), 133–138.

Agglomeration and the spatial determinants of productivity and trade

Long Abstract:

The tendency of human and economic activities to agglomerate is obvious, as proved by the existence of cities. This fact is also true for individual industries. The measure, causes, and consequences of the spatial concentration of firms operating in the same industry are an old subject of interest for economists, dating back at least to Alfred Marshall (1890).

These patterns of agglomeration cannot be explained by the existence of local comparative advantages only, as in the classical trade theory. It has led economists to acknowledge the existence of agglomeration economies. Such economies exist as soon as an individual's productivity rises when he or she is close to other individuals. Several questions are thus of interest: is spatial concentration pervasive in all industries, or limited to a few anecdotal cases? What are the advantages for firms to cluster that are able to offset extra costs due to agglomeration? How large are these advantages? How fast do they decline in space? How do they shape the spatial distribution of trade? These questions make up the background picture of this dissertation.

In chapter 1, we develop a new methodology to test for the spatial concentration of industries in a continuous space. Considering space as continuous prevents the measure of spatial concentration from being corrupted by statistical artifacts endemic to the use of a discrete spatial zoning system, the so-called Modifiable Areal Unit Problem. We emphasize in this chapter the specific patterns of concentration of service industries in comparison with the more traditional manufacturing industries in France. Service industries appear more localized than manufacturing industries and at shorter distances. This result supports the intuition that, in these industries, very localized agglomeration economies are crucial, as face-to-face contacts or off-the-street networking.

In chapters 2 and 3, we quantify the magnitude of agglomeration economies on the productivity of firms. More specifically, we assess the relative strength of urbanization and localization economies. In the former case, firms benefit from the overall size of their market, whatever the identity of their neighbors. In the latter case, firms benefit from the closeness of neighbors operating within the same industry. In chapter 2, we rely on a traditional linear-in-mean OLS approach and show that employment density and local specialization are of primary importance to explain disparities in average firm productivity over space. Firms located in areas of the upper decile for density are about 8% more productive than firms located in areas of the lower decile. This is equivalent to a four- to five-year growth in productivity. In comparison, firms in areas of the upper decile for specialization are, on average, 5% more productive than firms in areas of the lower decile.

In chapter 3, we use a quantile regression approach to assess whether these average results hide large differences across heterogeneous producers. Indeed, we find that urbanization economies benefit more the most productive firms. The productivity premium for firms located in areas of the upper decile for density ranges from 6% for the least productive firms to almost 14% for the most productive ones, in comparison with firms located in

areas of the lower decile for density. In comparison, the impact of specialization is rather stable across the conditional productivity distribution.

Chapter 4 studies another aspect of spatial concentration. We investigate whether and how the spatial distribution of immigrants across French départements shapes the international trade flows of these areas with the immigrants' countries of origin. We show that immigrants exert a positive impact on exports and imports. Doubling the number of immigrants settled in a *département* boosts its exports to the home country by 7% and its imports by 4%. This impact is larger when immigrants originate from a country with weak institutions or when the traded good is more complex. In both cases, it sustains the fact that immigrants hold specific knowledge about their country of origin that eases the creation of trade links.

Chapter 5 wraps this dissertation up by considering the sensitivity of the various econometric results shown in previous chapters to the Modifiable Areal Unit Problem. We study how changing the shape (equivalently, the drawing of their boundaries) and size (equivalently, the number) of spatial units impacts on the degree of spatial concentration, the magnitude of agglomeration economies and the spatial determinants of trade. We compare these distortions with those due to misspecification. This exercise is all the more important because most empirical work in regional and urban economics relies on scattered geocoded data that are aggregated into discrete spatial units, such as cities or regions. All of these empirical exercises suggest that, when spatial units remain small, changing their size only slightly alters economic geography estimates, and changing their shape matters even less. Both distortions are of secondary concern compared to specification issues.

Keywords: Spatial concentration, Agglomeration economies, Firm productivity, Immigration and Trade, Modifiable Areal Unit Problem.

Printed on June 30, 2010
Paris

Déterminants de la productivité et du commerce : le rôle de la proximité géographique

Cette thèse s'intéresse à la manière dont les économies d'agglomération façonnent l'organisation spatiale des secteurs d'activité en France et agissent sur la productivité des entreprises. Dans le premier chapitre, nous développons une nouvelle méthode pour mesurer la concentration spatiale à partir de données géo-localisées. Nous l'appliquons à la comparaison de la concentration spatiale dans les secteurs de service et dans les secteurs manufacturiers. Nous soulignons, entre autres, une tendance plus forte des secteurs de service à se concentrer spatialement. Dans les deuxième et troisième chapitres, nous évaluons l'impact des externalités d'urbanisation et de localisation sur la productivité des entreprises. Nous montrons dans le chapitre 2 que les entreprises gagnent, en moyenne, à être localisées dans une zone à forte densité en emploi et à proximité d'entreprises opérant dans le même secteur d'activité. Le chapitre 3 étudie comment ces effets sont différenciés entre entreprises hétérogènes. Nous soulignons le fait que les entreprises les plus productives sont aussi celles qui bénéficient le plus des externalités d'urbanisation. Le chapitre 4 se concentre sur les problématiques de commerce international. Nous montrons que plus le nombre d'immigrés est grand dans un département français, plus ses échanges commerciaux avec le pays d'origine de ces immigrés sont importants. Enfin, le dernier chapitre propose une contribution méthodologique. Nous étudions comment le choix particulier d'un découpage géographique, avec des unités spatiales de taille et de forme données, influence les résultats des exercices statistiques des chapitres précédents. Nous concluons à un biais faible lié au Problème des Unités Spatiales Modifiables, au regard du biais introduit par une mauvaise spécification.

Mots clés: Concentration spatiale, Économies d'agglomération, Productivité des entreprises, Immigration et commerce, Problème des Unités Spatiales Modifiables

Agglomeration and the spatial determinants of productivity and trade

This PhD dissertation studies how agglomeration economies shape the patterns of spatial concentration in French industries, and impact on French firm productivity. In the first chapter, we develop a new methodology to assess spatial concentration with micro-geographic data. This methodology is then applied to compare localization patterns in French service and manufacturing industries. In particular, we find that service industries tend to be more localized than manufacturing ones. In the second and third chapters, we assess the magnitude of urbanization and localization economies on French firm productivity. Chapter 2 proves that, on average, firms benefit from a larger density of employment in their vicinity and a more specialized environment. Chapter 3 considers the differential impact of agglomeration economies across heterogeneous producers. We emphasize that urbanization economies benefit more the most productive firms. Chapter 4 focuses on international trade issues. We find that the larger the stock of immigrants in a specific French département, the larger its trade flows toward the immigrants' country of origin. Finally, chapter 5 makes a methodological point by considering whether and how the choice of a specific zoning system, with spatial units of given size and shape, impacts on the statistical exercises of the previous chapters. We find that distortions due to the Modifiable Areal Unit Problem are of secondary concern in comparison with problems due to misspecification.

Keywords: Spatial concentration, Agglomeration economies, Firm productivity, Immigration and Trade, Modifiable Areal Unit Problem.